DOCUMENT RESUME

ED 095 211 TM 003 884

TITLE	New Trends in Evaluation. Report No. 35.
INSTITUTION	Gothenburg Univ. (Sweden). Inst. of Education.
SPONS AGENCY	National Swedish Board of Education, Stockholm.
REPORT NO	R-35
PUB DATE	Jan 74
NOTE	92p.; Parts in Swedish
EDRS PRICE	MF-$0.75 HC-$4.20 PLUS POSTAGE
DESCRIPTORS	*Educational Objectives; *Educational Researchers; *Ethics; Formative Evaluation; *Program Evaluation
IDENTIFIERS	*Research Methods

ABSTRACT

This report contains an introductory paper by Robert E. Stake, "Program Evaluation, Particularly Responsive Evaluation," to a conference, "New Trends in Evaluation," at the Institute of Education, Gothenburg University, Sweden. Also included are comments on State's paper by two invited discussants, Eskil Bjorklund and Terence Marton. It further contains summaries of the group discussions (in Swedish) together with lists of participants. Stake states that the choice of the general strategy and the particular methods for an evaluation study must depend on the questions to be answered, and for answering some kinds of questions, responsive evaluation is a better alternative. He mentions two major program evaluation models: the informal study and the pretest/posttest model, based on prespecified behavioral objectives. In comment, Marton states that responsive evaluation, can be seen as a denial of, an alternative to, or a special type of educational evaluation, and comments on its relationship to other areas of evaluation. (RC)

ED 095211

BEST COPY AVAILABLE

U.S. DEPARTMENT OF HEALTH
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Reports from

THE INSTITUTE OF EDUCATION
UNIVERSITY OF GÖTEBORG

NEW TRENDS IN EVALUATION

Report from a conference on educational evaluation with an introductory paper by professor Robert E. Stake, University of Illinois

No. 35
January 1973

ED 095211
IR 000 884

INSTITUTE OF EDUCATION · MÖLNDALSVÄGEN 24 · GÖTEBORG S · SWEDEN

ED 095211

NEW TRENDS IN EVALUATION

Report from a conference on educational evaluation with an introductory paper by professor Robert e. Stake, University of Illinois

TM 003 884

INNEHÅLLSFÖRTECKNING

PREFACE

PROGRAM EVALUATION, PARTICULARLY RESPONSIVE EVALUATION. 1
Robert E. Stake

COMMENTS ON ROBERT E. STAKE: "PROGRAM EVALUATION, PARTICULARLY RESPONSIVE EVALUATION." Eskil Björklund 21

COMMENTS ON ROBERT E. STAKE: "PROGRAM EVALUATION, PARTICULARLY RESPONSIVE EVALUATION." Ference Marton 26

SAMMANDRAG AV DISKUSSIONERNA I GRUPP A. (73.10.22-23) 34
Roger Säljö och Per Olof Thång

SAMMANDRAG AV DISKUSSIONERNA I GRUPP B. (73.10.22) 37
Lars Aldemark

SAMMANDRAG AV DISKUSSIONERNA I GRUPP B. (73.10.23) 43
Gunilla Svingby

SAMMANDRAG AV DISKUSSIONERNA I GRUPP C. (73.10.22) 47
Claes G. Wenestam

SAMMANDRAG AV DISKUSSIONERNA I GRUPP C. (73.10.23) 53
Lennart H.S. Nilsson

SAMMANDRAG AV DISKUSSIONERNA I GRUPP D. (73.10.22-23) 58
Marianne Colliander och Kerstin Rosén

SAMMANDRAG AV DISKUSSIONERNA I GRUPP E. (73.10.22) 64
Gunnel Johansson

SAMMANDRAG AV DISKUSSIONERNA I GRUPP E. (73.10.23) 68
Christina Gustafsson

SAMMANDRAG AV DISKUSSIONERNA I GRUPP F. (73.10.22-23) 73
Lars Eriksson och Lars Ove Dahlgren

SAMMANDRAG AV DISKUSSIONERNA I GRUPP G. (73.10.22-23) 79
Göran Patriksson och Ulla-Britt Bladini

DELTAGARLISTA 82

REFERENSER 85

PREFACE

During the second part of 1973 the Institute of Education at Göteborg University has as its guest Professor Robert E. Stake of the University of Illionois. In order to take advantage of his presence the Institute arranged in October 1973 a two-day conference on "NEW TRENDS IN EVALUATION". Educators and researchers in Sweden and the other Scandinavian countries were invited and a great number took this opportunity to meet and discuss writings and developments in the evaluation of educational programs.

This resport contains the introductory paper by Robert Stake and the comments by two invited discussants, Eskil Björklund of the Office of the Chancellor of Swedish Universities and Ference Marton of Göteborg University. It further contains summaries of the group discussions (for practical reasons in Swedish language) together with lists of participants.

The conference was supported by a grant from the Swedish National Board of Education.

Finally we wish to thank all who contributed to the successful proceedings of the conference.

Kjell Härnqvist Ulf Lundgren Erik Wallin

BEST COPY AVAILABLE

PROGRAM EVALUATION, PARTICULARLY RESPONSIVE EVALUATION

Robert E Stake
Center for Instructional Research and Curriculum Evaluation
University of Illinois

I am pleased to have this opportunity to talk about some recent developments in the methodology of program evaluation, and about what I call "responsive evaluation".

I feel fortunate to have, not only these two days, but some seven months to think about these things. My hosts here at the Gothenburg Institute of Educational Research have been most hospitable, but generous also in hearing me out, pointing my head in still another way, weighing the merit of our several notions, and offering occasionally the luxury of a passionate argument.

When Erik or Hans or Sverker or Ulf and I agree, we are struck by the fact that the world is but one world and the problems of educationare universal. When we disagree they are quick to suggest that the peculiar conditions of education in America have caused me to make peculiar assumptions, and perhaps even warped my powers of reasoning. I am sure that some of you here today will share those findings. What I have to say is not only that we in educational research need to be doing some things we have not been doing, but that in doing what we have been doing we are in fact part of the problem.

Our main attention will be on program evaluation. A program may be strictly or loosely defined. It might be as large as all the teacher-training in the United States or it might be as small as a field-trip for the pupils of one classroom. The evaluation circumstances will be these: that someone is commissioned in some way to evaluate a program, probably an on-going program; that he has some clients or audiences to be of assistance to--usually including the educators responsible for the program; and that he has the responsibility for preparing communications with these audiences.

In 1965 Lee Cronbach, then President of the American Educational Research Association, asked me to chair a committee to prepare a set of standards for evaluation studies, perhaps like the Standards for Educational and Psychological Tests and Manuals compiled by John French and Bill Michael and published in 1966 by the American Psychological Association. Lee Cronbach, Bob Heath, Tom Hastings, Hulda Grobman, and other educational researchers have worked with many of the U.S. curriculum reform projects, in the 1950's and early 1960's, and had recognized the difficulty of evaluating curricula, and the great need for guidance on the design of evaluation studies.

Our committee reported that it was too early to decide upon a particular method or set of criteria for evaluating educational programs, that what educational researchers needed was a period of field work and discussion to gain more experience in how evaluative studies could be done. Ben Bloom, successor to Lee Cronbach in the Presidency of AERA, got the AERA to sponsor a Monograph Series on Curriculum Evaluation for the purpose we recommended. The seven volumes completed under AERA sponsorship are shown on the handout sheet. The series in effect will continue under sponsorship of the UCLA Center for the Study of Evaluation, whose director, Marv Alkin, was a guest professor here at this Institute for Educational Research two years ago. I think this Monograph Series can take a good share of the credit, or blame, for the fact that by my count over 200 sessions at the 1973 AERA annual meeting programs were directly related to the methods and results of program evaluation studies.

There were two primary models for program evaluation in 1965, and there are two today. One is the informal study, perhaps a self-study, usually using information already available, relying on the insights of professional persons and respected authorities. It is the approach of regional accrediting associations for secondary schools and colleges in the United States, and is exemplified by the Flexner report (1916) of medical education in the USA and by the Coleman report (1966) of equality of educational opportunity. On the sheet you received with your background reading materials, one entitled Prototypes of Curriculum Evaluation, I have ever-so briefly described this and other models; this one is referred to there as the School Accreditation Model. Most Educators are partial to this evaluation model, moreso if they can specify who the panel members or examiners are. Researchers do not like it because it relies so much on second-hand information. But there is much good about the model.

Most researchers have preferred the other model, the pretest/post-test model, what I have referred to on the prototype sheet as Ralph Tyler's model. It often uses prespecified statements of behavioral objectives--such as are available from Jim Popham's Instructional Objectives Exchange--and is nicely represented by Tyler's Eight Year Study, Husén's International Education Study and the National Asessement of Educational Progress. The focus of attention with this model is primarily on student performance.

Several of us have proposed other models. In a 1963 article is Cronbach's preference to have evaluation studies considered applied research on instruction, to learn what could be learned in general about curriculum development, as was done on Hilda Taba's Social Studies Curriculum Project. Mike Scriven strongly criticized Cronbach's choice in AERA Monograph No I, stating that it was time to give consumers (purchasing agents, taxpayers, and parents) information on how good each existing curriculum is. To this end, Kenneth Komoski established in New York City an Educational Product Information Exchange, which has reviewed equipment, books and teaching aids, but has to this day still not caught the buyer's eye.

Dan Stufflebeam was one who recognized that the designs preferred by researchers did not focus on the variables that educational administrators have control over. With support from Egon Guba, Dave Clark, Bill Gephart and others, he proposed a model for evaluation that emphasized the particular decisions that a program manager will face. Data-gathering would include data on Context, Input, Process and Product, but analysis would relate those things to the immediate management of the program. Though Mike Scriven criticized this design too, saying that it had too much bias toward the concerns and the values of the education establishment, this Stufflebeam CIPP model was popular in the U.S. Office of Education for several years. Gradually it fell into disfavor because it was not generating the information--or the protection--that program sponsors and directors needed. But that occured, I think, not because it was a bad model, but partly because managers were unable or unwilling to examine their own operations as part of the evaluation. Actually no evaluation model could have succeeded. A major obstacle was a federal directive which said that no federal office could spend its funds to evaluate its own work, that that could only be done by an office higher up.--Perhaps the best examples of evaluation reports following this approach are those done in the Pittsburgh schools by Mal Provus and Esther Kresh.

Before I describe the approach that I have been working on--which I hope will someday challenge the two major models--I will mention several relatively recent developments in the evaluation business.

It is recognized, particularly by Mike Scriven and Ernie House, that co-option is a problem, that the rewards to an evaluator for producing a favorable evaluation report often greatly outweigh the rewards for producing an unfavorable report. I do not know of any evaluators who falsify their reports, but I do know many who consciously or unconsciously choose to emphasize the objectives of the program staff and to concentrate on the issues and variables most likely to show where the program is successful. I often do this myself. Thus the matter of "meta-evaluation", providing a quality-control for the evaluation activities, has become an increasing concern.

Early in his first term of office President Nixon created a modest Experimental Schools Program, a program of five-year funding for three carefully selected high schools, (from all those in the whole country) and the elementary schools that feed students into them. Three more have been chosen each year, according to their proposal to take advantage of a broad array of knowledge and technical developments, and to show how good a good school can be. The evaluation responsibility was designed to be allocated at three separate levels, one internal at the local school level, one external at the local school level (i.e., in the community, attending to the working of the local school, but not controlled by it) and a third at the national level, synthesizing results from the local projects and evaluating the organization and effects of the Experimental Schools Program as a whole. Many obstacles and hostilities hampered the work of the first two evaluation teams. And work at the third level--according to Egon Guba who did a feasibility study--was seen to be so likely to fail that it probably should be carried no further.

Mike Scriven has made several suggestions for meta-evaluation, one most widely circulated based on abstinance, called "Goal-free evaluation". Sixten Marklund has jokingly called it "Aimless evaluation". But it is a serious notion, not to ignore all idea of goals, but to refrain completely from any personal discussion of goals with the program sponsors or staff. The evaluator perhaps with the help of colleagues and consultants, then is expected to recognize manifest goals and accomplishments of the program as he works it in the

field. Again, with the concern for the consumer of education, Scriven has argued that what is intended is not important, that the problem is a failure if its results are so subtle that they do not penetrate the awareness of an alert evaluator. Personally I fault Scriven for expecting us evaluators to be as sensitive, rational and alert as his designs for evaluation require. I sometimes think that Mike Scriven designs evaluation studies that perhaps only Mike Scriven is capable of carrying out.

Another interesting development is the use of adversarial procedures in obtaining evidence of program quality and especially in presenting it to decision makers. Tom Owens, Murray Levine, and Marilyn Kourilsky have taken the initiative here. They have drawn upon the work of legal theorists who claim that truth emerges when opposing forces submit their evidence to cross-examination directly before the eyes of judges and juries. Graig Gjerde, Terry Denny and I tried someting like this in our TCITY REPORT. You have a copy of it in the conference reading materials you received several weeks ago. If you have that orange-colored document with you you might turn to the very last pages, pages 26 and 27. On page 26 you find a summary of the most positive claims that might reasonably be made for the Institute we were evaluating. On page 27 is a summary of the most damaging charges that might reasonably be made. It was important to us to leave the issue unresolved, to let the reader decide which claim to accept, if any. But we would have served the reader better if we had each written a follow-up statement to challenge the other's claims. At any rate, this is an example of using an adversary technique in an evaluation study.

Now in the next 45 minutes or so I want to concentrate on the approach for evaluating educational programs presently advocated by Malcolm Parlett of the University of Edinborough, Barry McDonald of the University of East Anglia, Lou Smith of Washington University of St Louis, Bob Rippey of the University of Connecticut, and myself. You have had an opportunity to read an excellent statement by Malcolm Parlett and David Hamilton. Like they did, I want to empasize the settings where learning occurs, teaching transactions, judgement data, holistic reporting, and giving assistance to educators. I should not suggest that they endorse all I will say today, but their writings for the most part are harmonius with mine.

Let me start with a basic definition, one that I got from Mike Scriven. Evaluation is an

OBSERVED VALUE compared to some STANDARD. It is a simple ratio but this numerator is not simple. In program evaluation it pertains to the whole constellation of values held for the program. And the denominator is not simple, for it pertains to the complex of expectations and criteria that different people have for such a program.

The basic task for an evaluator is made barely tolerable by the fact that he does not have to solve this equation in some numerical wayror to obtain a descriptive summary grade, but needs merely to make a comprehensive statement of what the program is observed to be, with useful references to the satisfaction and dissatisfaction that appropriately selected people feel toward it. Any particular client may want more than this, but this satisfies the minimum concept, I think, of an evaluation study.

If you look carefylly at the TCITY REPORT you will find no direct expression of this formula, but it is in fact the initial idea that guided us. The form of presentation we used was chosen to convey a message about the Twin City Institute to our readers in Minneapolis and St. Paul, rather than to be a literal manifestation of our theory of evaluation.

Our theory of evaluation emphasizes the distinction between a preordinate approach and a responsive approach. In the recent past the major distinction being made by methodologists is that between what Scriven called formative and summative evaluation. He gave attention to the difference between developing and already developed programs, and implicitly to evaluation for a local audience of a program in specific setting as contrasted to evaluation for many audiences of a potentially generalizable program. These are important distinctions, but I find it even more important to distinguish between preordinate evaluation studies and responsive evaluation studies.

I have made the point that there are many different ways to evaluate educational programs. No one way is the right way. Some highly recommended evaluation procedures do not yield a full description, nor a view of the merit and shortcoming of the program being evaluated. Some procedures ignore pervasive questions that should be raised whenever educational programs are evaluated:

- Do all students benefit or only a special few?
- Does the program adapt to instractors with unusual qualifications?
- Are opportunities for aesthetic experience realized?

Some evaluation procedures are insensitive to the uniqueness of the local conditions. Some are insensitive to the quality of the learning climate provided. Each way of evaluating leaves some things de-emphasized.

I prefer to work with evaluation designs that perform a service. I expect the evaluation study to be useful to specific persons. An evaluation probably will not be useful if the evaluator does not know the interests and language of his audiences. During an evaluation study, a substantial amount of time may be spent learning about the information needs of the persons for whom the evaluation is being done. The evaluator should have a good sense of whom he is working for and their concerns.

Responsive Evaluation

To be of service and to emphasize evaluation issues that are important for each particular program, I recommend the responsive evaluation approach. It is an approach that sacrifies some precision in measurement, hopefully to increase the usefulness of the findings to persons in and around the program. Many evaluation plans are more "preordinate", emphasizing (1) statement of goals, (2) use of objective tests, (3) standards held by program personnel, and (4) research-type reports. Responsive evaluation is less reliant on formal communication, more reliant on natural communication.

Responsive evaluation is an alternative, an old alternative. It is evaluation based on what people do naturally to evaluate things: they observe and react. The approach is not new. But it has been avoided in planning documents and institutional regulations because, I believe, it is subjective, poorly suited to formal contracts, and a little too likely to raise the more embarassing questions. I think we can overcome the worst aspects of subjectivity, at least. Subjectivity can be reduced by replication and operational definition of ambigious terms, even while we are relying heavily on the insights of personal observation.

An educational evaluation is responsive evaluation (1) if it orients more directly to program activities than to program intents, (2) if it responds to audience requirements for information, and (3) if the different value-perspectives of the people at hand are referred to in reporting the success and failure of the program. In these three separate ways an evaluation plan can be responsive.

To do a responsive evaluation, the evaluator of course does many things. He makes a plan of observations and negotiations. He arranges for various persons to observe the program. With their help he prepares brief narratives, portrayals, product displays, graphs, etc. He finds out what is of value to his audiences. He gathers expressions of worth from various individuals whose points of view differ. Of course, he checks the quality of his records. He gets program personnel to react to the accuracy of his portrayals. He gets authority figures to react to the importance of various findings. He gets audience members to react to the relevance of his findings. He does much of this informally, literating, and keeping a record of action and reaction. He chooses media accessible to his audiences to increase the liklihood and fidelity of communication. He might prepare a final written report, he might not--depending on what he and his clients have agreed on.

Purposes and Criteria

Many of you will agree that the book edited by E.F. Lindquist, Educational Measurement, has been the bible for us who have specialized in educational measurement. Published in 1951 it contained no materials on program evaluation. The second edition, edited by Bob Thorndike, has a chapter on program evaluation. Unfortunately, the authors of this chapter, Alex Astin and Bob Panos, chose to emphasize but one of the many purposes of evaluation studies. They said:

> "The principal purpose of evaulation is to produce information that can guide decisions concerning the adoption or modification of an educational program."

People expect evaluation to accomplish many different purposes:

- to document events
- to record student change

- to detect institutional vitality
- to place the blame for trouble
- to aid administrative decision making
- to facilitate corrective action
- to increase our understanding of teaching and learning

Each of these purposes is related directly or indirectly to the values of a program, and may be a legitimate purpose for a particular evaluation study. It is very important to realize that each purpose needs separate data, all the purposes cannot be served with a single collection of data. Only a few questions can be given prime attention. We should not let Astin and Panos decide what questions to attend to, or Tyler, or Stake. Each evaluator, in each situation, has to decide what to attend to. The evaluator has to decide.

On what basis will he choose the prime questions? Will he rely on his preconceptions? Or on the formal plans and objectives of the program? Or on actual program activities? Or on the reactions of participants? It is at this choosing that an evaluator himself is tested.

Most evaluators can be faulted for over-reliance on preconceived notions of success. I advise the evaluator to give careful attention to the reasons the evaluation was commissioned, then to pay attention to what is happening in the program, then to choose the value questions and criteria. He should not fail to discover the best and worst of program happenings. He should not let a list of objectives or an early choice of data-gathering instruments draw attention away from the things that most concern the people involved.

Many of my fellow evaluators are committed to the idea that good education results in measurable outcomes: student performance, mastery, ability, attitude. But I believe it is not always best to think of the instrumental value of education as a basis for evaluating it. The "payoff" may be diffuse, long delayed; or it may be ever beyond the scrutiny of evaluators. In art education for example it is sometimes the purpose of the program staff

or parent to provide artistic experiences--and training--for the intrinsic value alone. "We do these things because they are good things to do", says a ballet teacher. Some science professors speak similarly about the experiential value of reconstructing certain classical experiments. The evaluator or his observers should note whether or not those learning experiences were well arranged. They should find out what appropriately selected people think are the "costs" and "benefits" of these experiences in the dance studio or biology laboratory. The evaluator should not presume that only measurable outcomes testify to the worth of the program.

Sometimes it will be important for the evaluator to do his best to measure student outcomes, other times not. I believe that there are few "critical" data in any study, just as there are few "critical" components in any learning experience. The learner is capable of using many pathways, many tasks, to gain his measure of skill and aesthetic "benefit". The evaluator can take different pathways to reveal program benefit. Tests and other data-gathering should not be seen as essential; neither should they be automatically ruled out. The choice of these instruments in responsive evaluation should be made as a result of observing the program in action and of discovering the purposes important to the various groups having an interest in the program.

Responsive evaluations require planning and structure; but they rely little on formal statements and abstract representations, e.g., flow charts, test scores. Statements of objectives, hypotheses, test batteries, teaching syllabi are, of course, given primary attention if they are primary components of the instructional program. Then they are treated not as the basis for the evaluation plan but as components of the instructional plan. These components are to be evaluated just as other components are. The proper amount of structure for responsive evaluation depends on the program and persons involved.

Substantive Structure

Instead of objectives, or hypotheses as "advanced organizers" for an evaluation study, I prefer issues. I think the word "issues" better reflects a sense of complexity, immediacy, and valuing. After getting acquainted with a program, partly by talking with students, parents, taxpayers, program sponsors, and program staff, the evaluator acknowledges

certain issues or problems or potential problems. These issues are a structure for continuing discussions with clients, staff, and audiences. These issues are a structure for the data-gathering plan. The systematic observations to be made, the interviews and tests to be given, if any, should be those that contribute to understanding or resolving the issues identified.

In evaluating TCITY, Craig Gjerde and I became aware of such issue-questions as:

- Is the admissions policy satisfactory?
- Are some teachers too "permissive"?
- Why do so few students stay for the afternoon?
- Is opportunity for training younger teachers well used?
- Is this Institute a "lighthouse" for regular school curriculum innovation?

The importance of such questions varies during the evaluation period. Issues which are identified early as being important tend to be given too much attention in a preordinate data plan, and issues identified toward the end are likely to be ignored. Responsive-evaluation procedures allow the evaluator to respond to emerging issues as well as to preconceived issues.

The evaluator usually needs more structure than a set of questions to help him decide "what data to gather". To help the evaluator conceptualize his "shopping list", I once wrote a paper entitled "The contenance of Educational Evaluation". It contained the matrix, the 13 information categories, shown in this presentation on the screen. You may notice that my categories are not very different from those called for in the models of Dan Stufflebeam and Mal Provus.

For different evaluation purposes there will be different emphases on one side of the matrix or the other: descriptive data and jurigemental data. And, similarly, there will be different emphases on antecedent, transaction, and outcome information. The "countenance" article also emphasized the use of multiple and even contradicting sources of information.

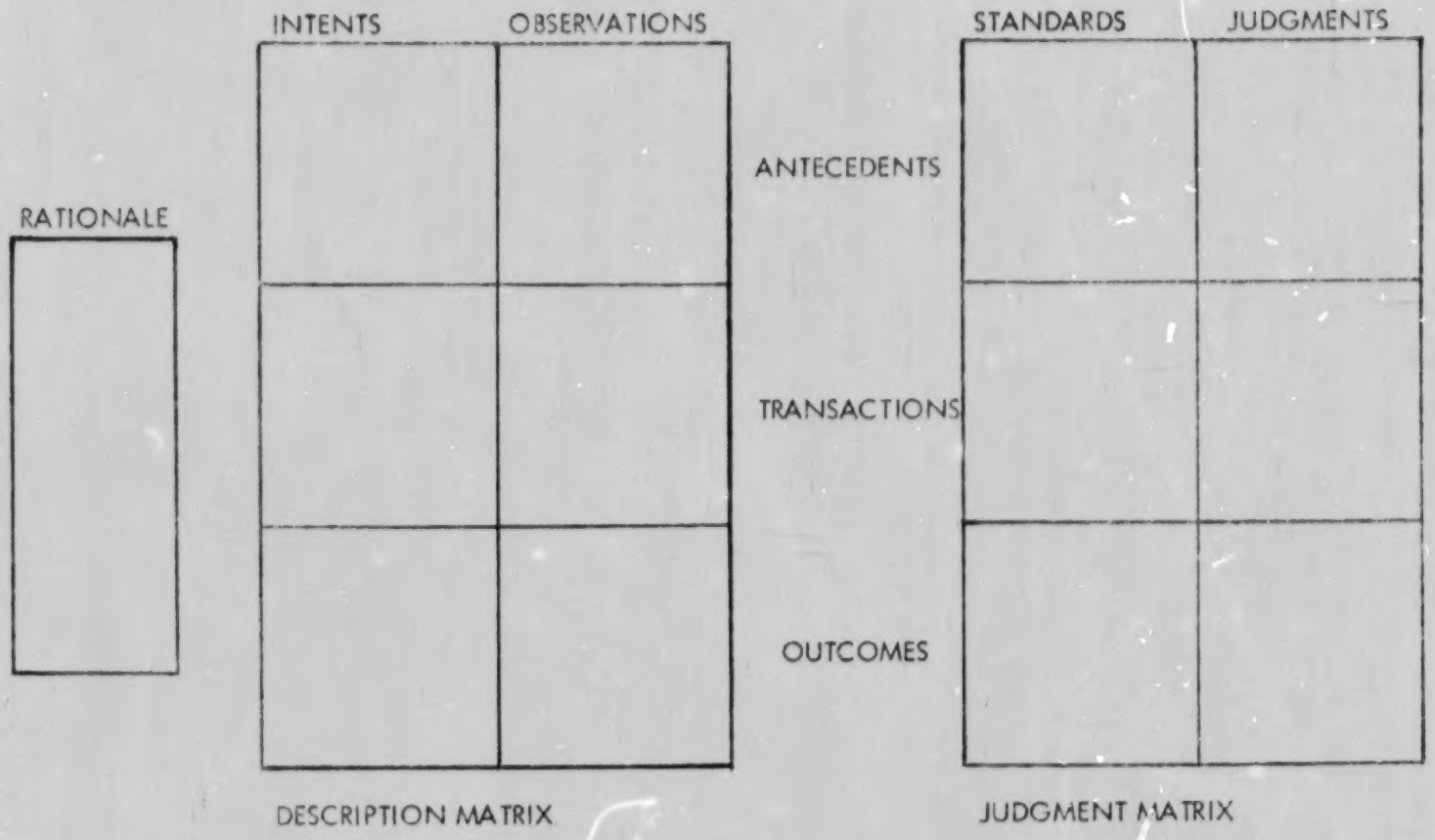


Figure 1. A layout of statements and data to be collected by the evaluator of an educational program.

Figure 2. Prominent events in a responsive evaluation.

It also pointed out the often ignored question about the match-up between intended instruction and observed instruction; and the even more elusive question about the strength of the contingency of observed outcomes upon observed transactions, under the particular conditions observed. I think these "countenance" ideas continue to be good ones for planning the content of the evaluation study.

I like to think of all of these data as observations: intents, standards, judgements, and statements of rationale are observed data too. Maybe it was a mistake to label just the second column "Observations". Thoreau said:

> Could a greater miracle take place than for us
> to look through each other's eyes for an instant.

Human observers are the best instruments we have for many evaluation issues. Performance data and preference data can be psychometrically scaled when objectively quantified data are called for. The important matter for the evaluator is to get his information in sufficient amount from numerous independent and credible sources so that it effectively represents the perceived status of the program, however complex.

Functional Structure

"Which data" is one thing but "how to do the evaluation" is another. My responsive-evaluation plan allocates a large expenditure of evaluation resources to observing the program. The plan is not divided into phases because observation and feedback continue to be the important functions from the first week trough the last. I have identified twelve recuring events. On the screen here I show them as if on the face of a clock. I know some of you would remind me that a clock moves clockwise, so I hurry to say that this clock moves clockwise and counter-clockwise and cross-clockwise. In other words, any event can follow any event. Furthermore, many events occur simultaneously, and the evaluator returns to each event many times before the evaluation ends.

For example, take twelve o'clock. The evaluator will discuss many things on many occasions with the program staff and with people who are representative of his audiences. He

will want to check his ideas of program scope, activities, purposes, and issues against theirs. He will want to show them his represenations (e.g. sketches, displays, portrayals, photographs, tapes) of value questions, activities, curricular content, and student products. Reactions to these representations will help him learn how to comminicate in this setting. He should provide useful information. He should not pander to desires for only favorable (or only unfavorable) information, nor should he suppose that only the concerns of evaluators and external authorities are worthy of discussion. (Of course, these admonitions are appropriate for responsive evaluation and preordinate evaluation alike).

This behavior of the responsive evaluator is very different from the behavior of the preordinate evaluator. Here on the screen now is my estimate as to how the two evaluators would typically spend their time.

	Preordinate	Responsive
Identifying issues, goals	10 %	10 %
Preparing instruments	30 %	15 %
Observing the program	5 %	30 %
Administering tests, etc.	10 %	-
Gathering judgments	-	15 %
Learning client needs, etc.	-	5 %
Processing formal data	25 %	5 %
Preparing informal reports	-	10 %
Preparing formal reports	20 %	10 %

I believe the preordinate evaluator conceptualizes himself as a stimulus, seldom as a response. He does his best to generate standardized stimuli, such as behavioral objective statements, test items, or questionnaire items. The responses that he evokes are what he collects as the substance of his evaluation report.

The responsive evaluator considers the principal stimuli to be those naturally occuring in the program, including responses of students and the subsequent dialogues. At first his job is to record these, learning both of happenings and values. For additional information he

assumes a more interventionist role. And with his clients and audience he assumes a still more active role, stimulating their thought (we hope) and adding to their experience with his reports.

Philosopher David Hawkins responded to the idea of reversing S - R roles in this way:

> ...I like the observation that one is reversing the S and R of it. In an experiment one puts the system in a prepared state, and then observes the behavior of it. Preparation is what psychologists call "stimulus", ... In naturalistic investigation one does not prepare the system, but looks for patterns, structures, significant evens, as they appear under conditions not controlled or modified by the investigator, who is himself now a system of interest. He is a resonator, a respondent. He must be in such an initial state that (a) his responses contain important information about the complex of stimuli he is responding to, and (b) they must be maximally decodable by his intended audience.

In the next section of this paper I will talk about maximally decodable reports. Let me conclude these two sections on structure by saying that the evaluator should not rely only on his own powers of observation, judgment and responding. He should enlist a platoon of students, teachers, community leaders, curriculum specialists, etc.,--his choice depending on the issues to be studied and the audiences to be served. The importance of their information, and the reliability of it, will increase as the number and variety of observers increase.

Portrayal and Holistic Communication

Maximally decodable reports require a technology of reporting that we educational measurements people. We have tried to be impersonal, theoretical, generalizable. We have sought the parsimonious explanation. We have not accepted the responsibility for writing in a way that is maximally comprehensible to practicing educators and others concerned about education. According to R.F. Rhyne:

> There is a great and growing need for the kind of powers of communication that help a person gain, vicariously, a feeling for the natures of fields too extensive and diverse to be directly experienced.

> Prose and its archetype, the mathematical equation, do not suffice. They offer more specificity within a sharply limited region of discourse than is safe, since the clearly explicit can be so easily mistaken for truth, and the difference can be large when context is slighted.

We need this power of communication, this opportunity for vicarious experience, in our attempts to solve educational problems.

One of the principal reasons for backing away from the preordinate approach to evaluation is to improve communication with audiences. The conventional style of research-reporting is a "clearly explicit" way of communicating. In a typical research project, the report is limitied by the project design. A small number of variables are identified and relationships among them are sought. Individuals are observed, found to differ, and distributions of scores are displayed. Covariations of various kinds are analyzed and interpreted. From a report of such analytic inquiry it is very hard, often impossible, for a reader to know "what the program was like". If he is supposed to learn "what the program was like", the evaluation report should be different from the conventional research report.

As a part of my advocacy of the responsive approach I have urged my fellow evaluators to respond to what I believe are the natural ways in which people assimilate information and arrive at understanding. Direct personal experience is an efficient, comprehensive, and satisfying way of creating understanding[15], but a way not usually available to our evaluation-report audiences. The best substitute for direct experinece probably is vicarious experience--increasingly better when the evaluator uses "attending" and "conceptualizing" styles similar to those which members of the audience use. Such styles are not likely to be those of the specialist in measurement or theoretically minded social scientist. Vicarious experience often will be conceptualized in terms of persons, places, and events.

We need a reporting procedure for facilitating vicarious experience. And it is available. Among the better evangelists, anthropologists, and dramatists are those who are developed

the art of story-telling. We need to portray complexity. We need to convey holistic impression, the mood, even the mystery of the experience. The program staff or people in the community may be "uncertain". The audiences should feel that uncertainty. More ambiguity rather than less may be needed in our reports. Oversimplification obfuscates. Ionesco[16] said,

> As our knowledge becomes separated from life, our culture no longer contains ourselves (or only an insignificant part of ourselves), for it forms a "social" context into which we are not integrated.
>
> So the problem becomes that of bringing our life back into contact with our culture, making it a living culture once again. To achieve this, we shall first have to kill "the respect for what is written down in black and white ..." to break up our language so that it can be put together again in order to re-establish contact with "the absolute", or as I should prefer to say, with "multiple reality"; it is imperative to "push human beings again towards seeing themselves as they really are". (p. 298).

Some evaluation reports should reveal the "multiple reality" of an educational experience.

The responsive evaluator will often use portrayals. Some will be short, featuring perhaps a five-minute "script", a log, or scrapbook. A longer portrayal may require several media: narratives, maps and graphs, exhibits, taped conversations, photographs, even audience role-playing. Which ingredients best convey the sense of the program to a particular audience? The ingredients are determined by the structure chosen by the evaluator.

> Suppose that a junior-high-school art program is to be evaluated. For portrayal of at least one issue, "how the program affects every student", the students might be thought of as being in two groups: those taking at least one fine-arts course and those taking none. (The purpose here is description, not comparison).
>
> A random sample of ten students from each group might be selected and twenty small case studies developed. The prose description of what each does in classes of various kinds (including any involvement with the arts in school) might be supplemented with such things

as (1) excerpts from taped interviews with the youngster, his friends, his teachers, and his parents; (2) art products (or photographs, news clippings, etc., of same) made by him in or out of class; (3) charts of his use of leisure time; and (4) test scores of his attitudes toward the arts. A display (for each student) might be set up in the gymnasium which could be examined reasonably thoroughly in 10-20 minutes.

Other materials, including the plan, program, and staffing for the school, could be provided. Careful attention would be directed toward finding out how the description of these individual youngsters reveals what the school and other sources of art experience are providing in the way of art education.

It will sometimes be the case that reporting on the quality of education will require a "two-stage" communication. Some audiences will not be able to take part in such a vicarious experience as that arranged in the example above. A surrogate audience may be selected. The evaluator will present his portrayals to them; then he will question them about the apparent activity, accomplishments, issues, strengths and shortcomings of the program. He will report their reactions, along with a more conventional description of the program, to the true audiences.

These twenty displays could be examined by people specially invited to review and respond to them. The reviewers[17] might be students, teachers, art curriculum specialists, and patrons of the arts. They might also visit regular school activities, but most attention would be to the displays. These reviewers should be asked to answer such questions as "Based on these case studies, is the school doing its share of providing good quality art experience for all the young people?" and "Is there too much emphasis on disciplined creative performance and not enough on sharing the arts in ways that suit each student's own tastes?" Their response to these portrayals and questions would be a major part of the evaluation report.

The portrayal will usually feature descriptions of persons. The evaluator will find that case studies of several students may more interestingly and faithfully represent the educational program than a few measurements on all of the students.[18] The promise of gain is twofold: the readers will comprehend the total program, and some of the important complexity

of the program will be preserved. The several students usually cannot be considered a satisfactory representation of the many--a sampling error is present. The protests about the sampling error will be loud; but the size of the error may be small, and it will often be a satisfactory price to pay for the improvement in communication.

There will continue to be many research inquiries needing social survey technology and exact specification of objectives. The work of John Tukey, Torsten Husén, Ralph Tyler, Ben Bloom, and James Popham will continue to serve as a model for such studies.

Often the best strategy will be to select achievement test, performance tests, or observation checklists to provide evidence that prespecified goals were or were not achieved. The investigator should remember that such a preordinate approach depends on a capability to state the important purposes of education and a capability to discern the accomplishment of those purposes, and those capabilities sometimes are not at our command. The preordinate approach usually is not sensitive to ongoing changes in program purpose, nor to unique ways in which students benefit from contact with teachers and other learners, nor to dissimilar viewpoints that people have as to what is good and bad.

Elliot Eisner nicely summerized these insensitivities in AERA Monograph 3. He advocated consideration of expressive objectives --toward outcomes that are idiosyncratic for each learner and that are conceptualized and evaluated after the instructional experience; after a product, an awareness, or a feeling has become manifest, at a time when the teacher and learner can reflect upon what has occured. Eisner implied that sometimes it would be preferable to evaluate the quality of the opportunity to learn--the "intrinsic" merit of the experience rather than the more elusive "payoff" to use Scriven's terms.

In my own writing on evaluation I have been influenced by Eisner and Scriven and others who have been dissatisfied with contemporary testing. We see too little good measurement of complex achievements, development of personal styles and sensitivities. I have argued that few, if any, specific learning steps are truly essential for subsequent success in any of life's endeavors. I have argued that students, teachers, and other purposively selected observers exercise the most relevant critical judgements, whether or not their criteria are

in any way explicit. I have argued also that the alleviation of instructional problems is most likely to be accomplished by the people most directly experiencing the problem, with aid and comfort perhaps (but not with specific solutions or replacement programs) from consultants or external authorities. I use these arguments as assumptions for what I call the responsive evaluation approach.

Utility and Legitimacy

The task of evaluating an educational program might be said to be impossible if it were necessary to express verbally its purposes or accomplishments. Fortunately, it is not necessary to be explicit about aim, scope, or probable cause in order to indicate worth. Explication will usually make the evaluation more useful; but it also increases the danger of misstatement of aim, scope, and probable cause.

To layman and professional alike, evaluation means that someone will report on the program's merits and shortcomings. The evaluator reports that a program is "coherent", "stimulating", "parochial", and "costly". These descriptive terms are also value-judgement terms. An evaluation has occured. The validity of these judgements may be strong or weak; their utility may be great or little. But the evaluation was not at all dependent on a careful specification of the program's goals, activities or accomplishments. In planning and carrying out an evaluation study, the evaluator must decide how far to go beyond the bare bones ingredients: values and standards. Many times he will want to examine goals. Many times he will want to provide a portrayal from which audiences may form their own value judgements.

The purposes of the audiences are all important. What would they like to be able to do with the evaluation of the program? Chances are they do not have any plans for using it. They may doubt that the evaluation study will be of use to them. But charts and products and narratives and portrayals do affect people. With these devices persons become better aware of the program, develop a feeling for its vital forces, a sense of its disappointments and potential troubles. They may be better prepared to act on issues such as a change of enrollment or a reallocation of resources. They may be better able to protect the program.

Different styles of evaluation will serve different purposes. A highly subjective evaluation may be useful but not be seen as legitimate. Highly specific language, behavioral tasks, and performance scores are considered by some to be more legitimate. In America, however, there is seldom a greater legitimacy, than the endorsement of large numbers of audience-significant people. The evaluator may need to discover what legitimacies his audiences (and their audiences) honor. Responsive evaluation includes such inquiry.

Responsive evaluation will be particularly useful during formative evaluation when the staff needs help in monitoring the program, when no one is sure what problems will arise. It will be particularly useful in summative evaluation when audiences want an understanding of a program's activities, its strengths and shortcomings and when the evaluator feels that it is his responsibility to provide a vicarious experience.

Preordinate evaluation should be preferred to responsive evaluation when it is important to know if certain goals have been reached, if certain promises have been kept, and when predetermined hypotheses or issues are to be investigated. With greater focus and opportunity for preparation, preordinate measurements made can be expected to be more objective and reliable.

It is wrong to suppose that either a strict preordinate design or responsive design can be fixed upon an educational program to evaluate it. As the program moves in unique and unexpected ways, the evaluation efforts should be adapted to them, drawing from stability and prior experience where possible, stretching to new issues and challenges as needed.

COMMENT ON ROBERT E STAKE: "PROGRAM EVALUATION, PARTICULARLY RESPONSIVE EVALUATION"

Eskil Björklund

I want to comment on the context where program evaluation is placed - its importance depends partly on how it is related to other aspects of evaluation.

I see program evaluation as a special type of educational evaluation, and I want to comment on its relations to other areas of evaluation. As I see it, there are problems and developments in these other areas which tend to make program evaluation even more important in a total evaluation strategy than it has been up to now.

Tests, examination	Program evaluation	External follow-up
(knowledge definition)	(educational process definition)	(work experience etc. as criteria definitions)

External evaluation

Traditionally much of the interest in evaluation has been focused on external evaluation. In this type of evaluation job experience and job success is related to educational programs, and the problem is how to adjust content and activities in schools in order to develop capabilities needed in work situations.

This is of course a necessary and healthy tendency, but it implies risks for educational measurement, for using too crude measures, too crude definitions both for the educational product coming out of the program (performance, time), and for the work life criteria.

Examples of these difficulties can be found in research on education's productivity-increasing effects or on education's importance for shaping a more equal society. We had earlier in the 60.s very optimistic reports on the role of education, both as regards the individual and the society in general, both as regards economic effects (income etc.) and as regards social position.

Especially after the recent Jencks' report there is now in the USA a growing pessimism as regards education's possibilities to improve conditions. According to Jencks, educational changes are not, and cannot be, really efficient means for increasing equality, social or economical.

There are a number of objections against Jencks' ways to analyze and draw conclusions. At the same time it is hardly possible to deny that also with the exception of Jencks we have now a number of follow-up studies which question the possibilities for education as a reliable means for increasing both productivity and equality in the society.

It is, however, typical for studies of this kind - dealing with the long term effects of education - to work with large samples and with crude measures.

One consequence, which these developments should have for educational evaluators, it to focus more on the educational process, to be able to define and measure educational programs in a more sophisticated way. This could gradually lead to more correct and reliable measures on what education is and on its outcomes.

Education is now an activity which more than almost any other type of ectivity takes people's time and interest. It would therefore be reasonable that education - in the same way as work, family life, leasure time activities etc. - should be studied primarily on its own right, not as a means but as a process, as a social milieu, as an acticity, as a form for living and working.

The possibilities for having such a viewpoint accepted depend partly on which evaluation strategy is given the highest pricrity by the politicians. But their attitude to what is most important to measure as regards education will to a great deal depend on how well developed and how convincing methods for program evaluation that are available. This in turn, depends partly on where evaluators focus their interest when dealing with education.

It is therefore important that educational evaluators gradually focus their interest more on the educational activities, on education as a process rather than as a product.

Test construction

Another kind of argument for program evaluation comes from the way in which evaluation has been handled as regards tests, grades and examination.

In test construction the interest has, as we know, for a long time been focussed more on reliability than on validity, more on statistical treatment of results than on definitions of content. Here educational researchers and test constructors have - maybe more unintentionally than intentionally - contributed to a knowledge ideal which is characterized by fairly simple and shortlived factual knowledge. Also in this country we have, however, now some very promising tendencies in test construction, where, as a counterweight to this, test types are developed which measure more general knowledge abilities. The focus is here on comprehension of basic principles and thought structures. Such more basic knowledge can also be supposed to have a higher survival value than the typical memory knowledge.

This increased reliance on problem solving and understanding in the knowledge definition which is given in test items and examination tasks will have consequenses for actual forms of instruction, for the educational process.

Test forms and study forms are interdependent on each other. A striving to develop and use test types which express a knowlege ideal focussing on problem solving and understanding therefore presupposes a comparable development in study programs. This development in its turn depends on program evaluation: program development depends on program evaluation.

Responsive program evaluation

Stake has given a fairly detailed description of what he means by responsive evaluation. He has described what a responsive program evaluator does and what is the difference between his behavior and the behavior of the traditional evaluator. There are big differences both in work methods and not least in ways of reporting. A traditional evaluator reports in written form. Such a report is usually, we must perhaps admit, not very informative and not very controversial. A responsive evaluator tries harder to give a full description of the program, a more living description, journalistic, more than tradition-

ally scientific. He wants to reproduce what is most typical and important in the actual program, and he can probably sometimes also become rather controversial. Stake stresses education as an art and evaluation as an art. At the same time it is obvious that he sees responsive evaluation as a task for educational research. Stake wants educational research to take the responsibility for, to develop and make use of the methods that are needed for a good responsive program evaluation.

He also sees both the educational activity and the evaluator's reporting on that ectivity as a theatrical performance. This strong focussing on the unique single event really makes difficulties for educational research, which as all research wants and needs to be cumulative. How can one responsive evaluation report be compared with and added to another, what about the survival value in a typical responsive evaluation report?

This is a real challenge for educational research. Can educational research broaden its theoretical and methodological approches so much that responsive program evaluation is incorporated into the activites of the educational research institutes?

Another perhaps more technical question, which is rather important to me, has to do with the role of the evaluator in responsive program evaluation. Stake stresses the evaluator's role as an expert coming in from the outside, as a consultant. To me it seems very natural, especially when dealing with higher education, to see the role of the program director, the study rector, the teaching team, as including the responsibility of program evaluation.

I am here thinking of, for example, univeristy institutes, where a number of course programs are run by groups of teachers and students. Would it not be reasonable in such circumstances to see program evaluation as an built-in activity rather than as a study to be commissioned to evaluation specialists coming in from outside? I hope this question can be discussed in the group work.

Program evaluation in Sweden

I will finish these general remarks by expressing the hope that this conference will increase the interest in program evaluation in this country. It has traditionally been relatively small.

Stake refers in his paper to Lindquist: "Educational measurement", 1951, that it contained no special part dealing with program evaluation.

I looked into the newest Scandinavian handbook on university teaching, published in 1973. It treats evaluation rather lengthy, but it does not deal with program evaluation.

The handbook has one chapter on the dimensions of evaluation, but this discussion is focussed on grading and final examination (where some of the procedures now being used correctly are critizised).

There is also a more how-to-do-it chapter on evaluation methods. This chapter deals with test construction, "tests being the by far most usual evaluation instrument". I would hope that within some years we can have a revised version of this handbook, with a substantial chapter on the various techniques and methods of program evaluation.

COMMENTS ON ROBERT E. STAKE: "PROGRAM EVALUATION, PARTICULARLY RESPONSIVE EVALUATION"

Ference Marton

Bob Stake's paper may be considered a fairly modest proposal; the two main issues being the following:

1. The choice of the general strategy and the particular methods for an evaluation study must depend on the questions we are trying to answer.

2. For answering some kinds of questions, responsive evaluation is a better alternative.

The first statement is a presupposition to the second. It means, for example, that if we want to answer the question whether an educational program has succeeded in achieving certain specified goals, we have to choose o n e evaluation strategy (for instance, the preordinate one) whereas if we want to answer the question "What does a program look like" we have to choose another (for instance, the responsive type of evaluation).

On this level of generality, the above statements may seem selfevident, but there are those who would claim that they are not.

Two major models of program evaluation are mentioned in the paper:

a) the informal study and

b) the pretest/post-test model, based on prespecified behavioral objectives.

Parlett o. Hamilton (1972) mention two major paradigms for educational research:

a) the classical or agricultural-botany paradigm, and

b) the social anthropology paradigm.

Bob Stake is speaking of evaluation, Parlett o. Hamilton of research. Without going too deeply into the question of what research *really* is and of whether evaluation *in fact* is some kind of research, we can state that there *is* a relationship between research and evaluation; both activities have the aim of producing knowledge (or information). The purpose of this knowledge is of course, usually narrower in evaluation studies than in so-called "pure" research (this is true in terms of the intended generalizability of findings, but as regards the complexity of the phenomena under investigation the relationship is, as a rule, reversed).

For a long time there has been only one accepted paradigm for scientific research in the Western world: the positivistic one, roughly characterized by the idea of theory-neutral observational language and of logical rules for going from observation to theory and vice versa. Science is supposed to be algorithmic, i.e. it is thought that there is a single method applicable to all possible cases.

In recent years, however, this paradigm has been seriously questioned. In the area of educational research the foremost expression of the movement toward a new meta-theoretical approach is to be found in the pages of the 71st Yearbook of NSSE under the editorship of Lawrence C. Thomas (Thomas, 1972).

I consider the paper presented today and the other works referred to, which are written in the same spirit, as expressions of the general trend toward a philosophical redirection in both basic and applied research. And, of course, to those, who hold that there is but one correct way of conducting research or evaluation the relativistic position (that is, that a method can be right or wrong only in relation to a given problem) must be something of a challange.

If you wish to introduce a new view on a subject (for example, evaluation) you can suggest that:

a) the earliger view is wrong and yours is the right one,

b) the earlier view is one of several possible views, of which yours is also one, or

c) the earlier view is a special case of your more general model.

In fact, I am not quite sure which of these strategies Bob Stake has chosen when arguing for a new approach to evaluation. If the earlier view includes the proposition that it is the only correct way of thinking, as in the case of the evaluation models inspired by the educational technology movement, in its most orthodox form, you have to abandon it, at least, partially, when you introduce a new model.

Apart from these philosophical considerations, we can say that the kind of proposal presented by Bob Stake clearly belongs to the second category, i.e. he is introducing an alternative, which is an additional model to the earlier ones as he says:

> "Responsive evaluation is an alternative."

But, in fact, he also says:

> "Statements of objectives, hypotheses, test batteries, teaching syllabi are, of course, given primary attention if they are primary components of the instructional program. Then they are treated not as the basis for the evaluation plan but as components of the instructional plan."

As far as I can see these last statements can be interpreted to mean that responsive evaluation is superordinate to preordinate evaluation since components of the second are included in the set of possible components of the first. (This superordinate relationship, also applies to informal evaluation, and for the same reasons.)

What I am arguing is that responsive evaluation, as presented here today, can be seen as a denial of, an alternative to or a superordinate model to other forms of evaluation. In the last mentioned case we have the general, but rather empty statement that the right way of carrying out evaluation depends on the kind of questions we want to answer in carrying out the evaluation study. This is not much of a guide, but it does have the most important implications for our attitudes to evaluation - and not only to evaluation. We have to realize that there is no general strategy in terms of a preconceived sequence of undertakings applicable to any evaluation study. Evaluation is not algorithmic, nor is science or human conduct in general.

There are, of course, restrictions and there are possible solutions. We can speak of what not to do and what possibly to do. We have bits of knowledge, we have what we may call "local" theories (with a limited scope of generality and complexity) and we have various methods which can be utilized. However, different instruments are to be used as parts which are functionally related to the whole. And the whole-qualities are to be found in every concrete situation seen from the angle of the problem itself.

I think that the most interesting aspect of responsive evaluation is just that it is n o t p r e o r d i n a t e. The main effects of an educational program are not necessarily the achievement of some educational objectives, stated in advance, and the intentions of the participants (students and teachers) are not necessarily identical with the objectives, all through the program.

What, from a theoretical point of view, is still more important, however, is that even if the expected effects remain the same on some level of generality (e.g. that the interest in a school-subject will increase), we often do not know in advance what concrete form the effects will take (e.g. the borrowing of more books, more talk about the subject out of class). There is no one-to-one relationship between the s i g n i f i c a t e (interest, knowledge, comprehension etc.) and the s i g n of it, i.e. the concrete behavior that we accept as evidence. Furthermore the sign-significate relationship is not contextually invariant, but is preserved only under certain circumstances.

If you want to be able to observe a n y possible outcome of an educational program or rather the most significant instead of the preconceived outcome, you have to be open-minded.

One way of being open-minded is to rely on some kind of direct personal experience. Reports on direct experiences can either be treated as data on personal experiences or reality or data on the reality that has been experienced.

In the first case we can work in an entirely "scientific" or even preordinate way, for instance by drawing samples of observers, who prepare reports on their experiences in

a holistic way, by drawing samples of cases to be described in a holistic way or by drawing samples of listeners (or members of a surrogate audience) who are going to react in a non-analytical way to the holistic descriptions. The observers and the audience than have to be open-minded on behalf of the evaluation specialist, so to speak, and the evaluation specialist can stand by as an "administrator of subjectivity". I have been thinking along the same lines myself (Marton 1970, Levin o. Marton 1971) when I have discussed alternative ways of describing goals and standards. For instance, I have argued, one could use a sample of items, prepared by a sample of teachers for illustrating the teacher's interpretation of the goals.

You can trace several notions of some kind of sampling design in the paper discussed here, but I do not really believe that what Bob is after here is a complicated sampling design. Rather it seems to me as if he was much more interested in the second way of treating reports on direct experiences, namely not as data, but as statements of how things are. And I think it is this point in the paper that I am most doubtful about.

The direct experiences of the evaluation specialist or the other observers vill, if reported in informal prose, facilitate "vicarious experiences". There is nothing wrong with this, but I do not quite believe that.

> "Direct personal experience is an efficient, comprehensive, and satisfying way of creating understanding." (p. 15).

What about all the statements which parents take to be true concerning child-rearing, or teachers concerning education?

On the other hand, as regards the skills of communicating direct experiences in order to promote "vicarious experiences", there are people, much better equipped for the job than the average educational researcher; for instance, as Bob Stake points out: "better evangelists, anthropologists and dramatists". Neither do I feel sure about the notion that you should entirely adjust your way of reporting to your audience; I believe in the evaluation specialist having a pedagogical mission too. I am convinced that educational research and evaluation can be both more professional and more responsive.

I have one final reflexion that I should like to make. Earlier on I made a considerable fuss about whether Bob is arguing for a more general, relativistic view of evaluation or if he is proposing a new, alternative model of evaluation. Obviously he is doing both, perhaps with more emphasis on the latter than on the former.

I think, however, that the most important issue is that the evaluation business (like human life in general) is non-algorithmic. Nobody can tell you in advance how to evaluate a n y possible educational program. Or as Bob Stake expresses this:

"Each evaluator, in each situation, has to decide what to attend to."

I think there must be something wrong when "responsive evaluation" is given as one kind of evaluation and "illuminative evaluation" as another. Rather, I should say, that all evaluation has to be responsive and illuminative.

REFERENCES

Levin, L. o. Marton, F. Provteori och provkonstruktion. Stockholm: Almqvist o. Wiksell, 1971.

Marton, F. Evaluering av evalueringar. In G. Handal (ed.) Universitetsstudier under debatt. Oslo: Universitetsförlaget, 1970. pp 11-25.

Parlett, M. o. Hamilton, D. Evaluation as illumination: A new approach to the study of innovatory programs. Centre for Research in the Educational Sciences. University of Edinburgh, 1972. (Mimeo).

Thomas, L.G. (ed.) Philosophical redirection of educational research. The Seventy-first Yearbook of the National Society for the Study of Education. Chicago: The University of Chicago Press, 1972.

GRUPP A:

Axelsson, Rune
Bäck, Per
Ekholm, Mats
Gustafsson, Bettil
Jansson, Bengt-Eric
Klingberg, Göte
Lindström, Berner
O Lövgren, Esse
Nordström, Anna-Lena
R_2 Säljö, Roger
R_1 Thång, Per Olof
Westrin, Per Anders

RESEARCH METHODS

Group A (Esse Lövgren)	Should the evaluator be reluctant to use unsolicited observations and casual opinions as data? Can we validate or corroborate such data?
Group E (Olof Magne)	Should the methods of evaluation differ for different subject-matters, for different agegroups, for different types of students and teachers?
Group F (Karl Gustaf Stukát)	Will classroom interaction analyses, e.g., Flanders, be useful in evaluation studies?
Group G (Lars-Gunnar Holmström)	What do ideas about responsive evaluation suggest for the training or retraining of educational researchers?

SAMMANDRAG AV DISKUSSIONERNA I GRUPP A (73.10.22-23)

Roger Säljö och Per Olof Thång

Inledningsvis vill vi som referenter påpeka att nedanstående sammanfattning knappast gör den livliga debatt som fördes i gruppen rättvisa. Väl medvetna om detta har syftet med vår sammanfattning i stället blivit att abstrahera några av de huvudtemata kring vilka diskussionen rörde sig. Argumentationens spontanitet och fullödighet går visserligen på detta sätt förlorad, men eftersom möjligheten att väl fullfölja referentsysslan tycks vara omvänt proportionell mot debattens intensitet så förefaller oss detta vara den enda möjliga utvägen.

Den första delen av debatten utgick från frågeställningen "Vad är nytt i Stakes angreppssätt?" Gruppen enades om att Stakes metod inte innehåller några direkta nyheter med avseende på tekniker och metoder i samband med evaluering. Stakes sätt att angripa problemen innehåller dock en förskjutning av intresset från traditionella evalueringskriterier till en mer holistisk syn på utvärderingsproblemen. "Responsive evaluation" tar sin utgångspunkt i "klienternas" upplevda behov och strävan att snabb evaluering skall kunna påverka den aktuella utbildningen eller undervisningen enligt en formativ modell.

Metoden implicerar ett betonande av den explorativa evalueringsfasen. Möjligheterna att integrera affektiva och sociala komponenter i utvärderingen förefaller större vilket gruppen fann positivt. Vi fann dock att "responsive evaluation" snarare är att betrakta som en ansats (approach) än som en metod för evaluering. Tillvägagångssättet medför en vidgad syn på begreppet evaluering, vilket i sig är positivt, men medger inte att direkta jämförelser med andra utvärderingsstrategier kan göras.

Andra delen av första gruppdiskussionen centrerades kring frågorna "Varför kommer denna typ av evalueringsstrategi nu?" och "Möjliga tillämpningar av densamma?" En viktig anledning till att evalueringsmodellen av Stakes typ aktualiseras nu, fann gruppen vara det faktum att man i dag i större utsträckning betonar affektiva och sociala mål inom ramen för skolans verksamhet. Ytterligare en bidragande faktor torde vara att man inom samhällsvetenskaperna nu alltmer har kommit att omfatta ett vetenskapsideal som tillåter

"mjukare data". Vad beträffar möjligheterna till tillämpning begränsas dessa av det faktum att utvärderingen torde bli dyrbar och tidsödande.

Under andra dagens grupparbetspass diskuterade grupp A problemen rörande "Research Methods". Inledningsvis vill vi påpeka att vi fann att Stakes metod inte har den precision och stringens som krävs för att det skall vara meningsfullt att dryfta dess konsekvenser för forskning och forskarutbildning. Ansatsen är dock som tidigare påpekats en "mjukare metod" och sålunda skulle man kunna uppfatta den som en orietering mot den hermeneutiska kunskapstraditionen.

Om man skall uppfatta evaluering och forskning som lika i det avseendet att de försöker öka vårt vetande och vår förmåga att förutsäga eller förklara varför ett fenomen ser ut som det gör, så uppfyller Stakes ansats inte dessa krav. Dennes intresse är i stället att vara provokativ och att fungera i enlighet med en "counselor-roll". En gruppmedlem framförde åsikten att svenska förhållanden skiljer sig ganska avsevärt från de amerikanska förhållanden, vilka Stake arbetar utifrån. Den "mjukare" ansats som förespråkas förekommer redan på skilda nivåer i det svenska skolsystemet. Tanken på "system-externa" evaluerare som enbart utvärderar "program-results" har dessutom aldrig vunnit gehör i Sverige.

Responsiv evaluering är en suggestiv term. Den innesluter en mängd "mjuka" metoder samtidigt som man inte förnekar värdet av andra mer konventionella. Denna typ av evaluering löper, enligt vår mening, risken att försöka gripa om så mycket i varje situation att vi har svårt att uppfatta den som en väl profilerad ansats i samband med evalueringsproblematiken. Det är möjligt att denna vår åsikt beror på att symposiet behandlade evaluering som en företeelse i sig, skild från ett ämnesinnehåll.

GRUPP B:

R_1	Aldemark, Lars
	Elison, Åke
	Graham, Peter
	Höghielm, Robert
	Kallós, Daniel
O	Levin, Lennart
	Silverberg, Bengt
	Solstad, Arne
	Svanberg, Gunilla
R_2	Svingby, Gunilla
	Ölund, Ulf

ETHICS AND EPISTEMOLOGY

Group B
(Lennart Levin)

Is there a sufficient relationship between what we know about teaching and what we know about learning to justify gathering student information (including test performance data) in order to correct or improve instruction?

Is it reasonable to believe that borrowings from the legal profession, e.g., rules of evidence, procedures of adversarial crossexamination, might be useful in the evaluation of educational programs?

How important is it to identify the values held by the evaluator? Should safeguards against bias, narrowness, or irrelevance be built into the designs?

SAMMANDRAG AV DISKUSSIONERNA I GRUPP B (73.10.22)

Lars Aldemark

Inledning

Grupperna hade tre förslag till diskussionsfrågor:

1) Vad är nytt ifråga om Bob Stakes "responsive evaluation"?

2) Varför är hans metod aktuell just nu?

3) Hur kan vi som pedagogiska forskare tillämpa metoden?

Gruppdiskussionen knöt an till dessa frågeställningar, men kan struktureras efter tre huvudlinjer. Den första huvudlinjen rörde Bob Stake som förnyare eller revisor: Kommer han verkligen med någonting nytt eller är det en gammal metod som han återupplivar och kontrasterar mot den moderna amerikanska forskningen? Linje två konfronterade Skolöverstyrelsen med fältet: Har SÖ och fältet samma verklighetsuppfattning? Den tredje linjen som kan urskiljas i gruppdiskussionen rörde målen för evalueringen: Har vi som forskare tillräckliga förutsättningar för att veta vad vi egentligen utvärderar?

Diskussionen

Daniel Kallós öppnade diskussionen med att påstå att det egentligen inte är så mycket nytt Bob Stake har att komma med på evalueringssidan. Hans viktigaste insats är att ställa upp en metod som kontrasterar mot nuvarande praxis; att visa på ett sätt att vinna förståelse och kommunikation, utan krav på vad vi vanligen menar med vetenskaplig precision. Upplever vi responsive evaluation som ett brott mot ett invant tankemönster, eller kan vi inkorporera modellen - helt eller som ett komplement - till de nuvarande metoderna, frågade DK.

- Är det inte bara så att vad Stake tar upp, är sådant som vi i dag diskuterar inom intervjutekniken, när vi ställer strukturerade mot ostrukturerade intervjuer? Jag uppfattar kritiken som riktad närmast mot de metoder som tagits fram i USA. Stake riktar sin kritik mot de utanförstående experter som kommer in i en skolsituation för att göra ut-

värderingar, menade BS. Han gör ungefär som studierektorn när denne utvärderar undervisningen mindre vetenskapligt, men närmare verkligheten.

DK såg BS inlägg som ett konkret exempel på hur man kan uppfatta samma sak på olika sätt.

- Jag tolkar BS, som att han ser detta som en fråga om olika tekniker, sade DK. För Bob Stake är det inte alls fråga om nya metoder utan snarare en helt annan modell. I Sverige kan vi antagligen acceptera att använda metoderna men använder inte dessa av olika skäl. Vilka avarter inom USA-forskningen reagerar BS egentligen emot? Du sade att Bob Stake arbetade närmare verkligheten och mindre vetenskapligt. Men är det inte verkligheten vi söker? Detta är kontroversiellt! Har vi kanske helt olika verklighetsuppfattningar?

- De avarter jag talar om är närmast reminicenser från diskussioner med amerikaner, svarade BS. Dessa gav ett intryck av evalueringen som en slags sport - eller en främmande fågel som kommer in i en skolmiljö och lämnar spår efter sig i form av obegripliga rapporter. När det gäller vetenskap menar jag att denna kanske kan arbeta mer indirekt - genom att studera ett led i undervisningen kan man skaffa sig kunskap om andra.

- Nu är det så att lärarna upplever både SÖ och Ingvar Karlsson, som sådana främmande fåglar, menade DK. Är det fråga om en reell skillnad i verklighetsuppfattning?

- Det framgår av andra artiklar av Bob Stake att han är i polemik med de traditionella evaluerarna, ansåg RH. Utvärderingarna måste komma ner och kommunicera med uppdragsgivarna. När det gäller SÖ, är det ju faktiskt så att det är en myndighet som består av lärare. Hur kan vi då ha så skilda synsätt? Här är det snarare fråga om brist på kommunikation!

- I så fall håller jag med Bob Stake, konstaterade DK. Kommunikationsproblem är ju just vad han vill försöka lösa. Lärarnas situation i klassrummet är en helt annan än den på SÖ. Man befinner sig på en annan nivå när det gäller t.ex. synen på vad som är problem. Eller är det inte så att problemen helt enkelt är annorlunda? SÖ och lärarna

har helt skilda verkligheter, genom att man sitter på olika beslutsnivåer.

GS frågade hur responsive evaluation står i relation till de mål man har för utbildningen. SÖ är ute efter att lösa generella problem, Stakes metod riktar in sig på en lägre nivå och på ett mer begränsat syfte. Stake talar närmast om nya utbildningsprogram, snarare än om effektivitet i undervisningen. Men är det samma typ av problem han vill evaluera?

- Det är inte så mycket fråga om att ställa olika metoder mot varandra, menade AS. Vi säger här att de olika metoderna kan ge liknande resultat. Vi måste skilja forskarrollen från de metoder forskaren använder. Stakes modell är vansklig att tillämpa på ett system. Vilken nivå är det vi vill gå in på och evaluera? Vilka forskarroller är det som går in?

- Stake ger i och för sig inte mycket hjälp, man hans metod skiljer sig utan tvekan från de traditionella. Det här är någonting nytt, ansåg ÅE. Vi har ju här i landet genomfört stora evalueringsprojekt, men vad har man egentligen nått fram till? När det gäller SÖ, så är de lärare som finns där inte representativa för folket på fältet.

- Vem rår egentligen för schismen mellan forskare/lärare och administrationen, frågade AS. Jo - vi själva som forskare, eftersom vi är så metodcentrerade. Skyll därför inte på SÖ utan diskutera i stället forskarens roll i undervisningen.

- Vi väljer våra experter ganska slumpmässigt, framhöll RH. Det är inga speciella typer av lärare, utan sådana som kan göra sig fria för uppdrag inom SÖ.

RH pekade också på de starka ekonomiska krafter inom läromedelsindustrin som verkade inom skolan. Förlagen propagerade för olika nya läromedel och om dessa blev misslyckade fick SÖ skulden. Dessutom tolkas uttalanden från enskilda, tunga, SÖ-anställda som representativa för hela SÖ av folk ute på fältet. Sådana faktorer påverkar relationerna, menade RH.

Därefter lämnades SÖ och myndighetens förhållande till fältet. Diskussionen knöts i stället till en fråga, som Ulf P. Lundgren tidigare tagit upp: ramarna för vetenskapen.

- I och för sig är det inte så mycket som är nytt i Stakes metod, ansåg UÖ. Det är den allra äldsta typen av evaluering - den som alla lärare dagligen utför.

- Det är ett dilemma när pedagoger kommer med nyheter, som i själva verket är gamla, framhöll DK. Varför upplever vi då dessa metoder som nyheter? Antagligen därför att man inte riktigt vet vad man håller på med. Stake gör någonting som skulle kunna vara nyttigt: han skapar en plattform för kritisk analys. Vad skiljer mig som forskare från Bob Stake? Varför upplever inte vi i Sverige någon konflikt?

- Stake tror att det går att beskriva vad som händer i ett undervisningsprogram, fortsatte DK. Han tror att den skolade pedagogen kan ge en rimlig beskrivning och det är i n t e likgiltigt för honom v e m som målar ett porträtt av verkligheten. Men - Bob Stake har inte gjort klart för sig var problemet ligger. Det gäller inte bara forskarrollen eller den nivå man ger sig in på - utan utifrån vilka förutsättningar vi evaluerar!

- En god utgångspunkt också för evaluering är att ha ett mål klart för sig, menade BS. Och ett lämpligt sådant är att skaffa fram beslutsunderlag. Vad tjänar evalueringen till? Här syftar den ena till generaliseringar och den andre till att hjälpa praktikerna.

- Ett annat problem är, menade RH, vilka möjligheter jag som läsare har att kontrollera vad som hänt i en evalueringssituation av Stakes modell. Det är också svårt att uttala sig om generaliserbarhet och reproducerbarhet.

- Om Stake skulle evaluera ett målstyrt program skulle målen ingå som en del tillsammans med miljö, elevrelationer etc, påpekade DK. Responsive evaluation påverkas av verkligheten. Alla utvärderingar syftar dessutom till att ge beslutsunderlag - även Stakes. Det är bara det att Stakes betraktelsesätt närmast är hermeneutiskt.

DK invände mot RH:s synpunkter och menade att Stakes löser sina problem genom att ställa krav på evaluerarna i form av värdepremisser. Frågan om generaliserbarhet blir en fråga om på vilken nivå vi reproducerar, ansåg DK vidare.

- Hur är det med möjligheterna att tillämpa Stakes idéer, frågade GS. Fungerar hans modell på våra egna projekt, till exempel?

- Vilken typ av kunskap är det över huvud taget vi inhämtar och förmedlar? Är det sådan kunskap som gör att vi kan justera eller korrigera en undervisningssituation eller process, frågade AS. I så fall - i förhållande till vad ska vi justera? Vad ska vår kunskap användas till? I dag saknar vi tillräckliga förutsättningar för att veta vad vi egentligen evaluerar.

- Syftet med evalueringen är ju det viktiga, framhöll GS. Kunskap är det vanligaste kriteriet, traditionellt i betydelsen ökade kunskaper och färdigheter i olika ämnen. Ett annat mål är att anpassa folk till samhället och i någon mån att skapa samhällsförändrare. Frågan är hur sådana mål ska utvärderas. Hur kan vi mäta om en elev blivit mer samarbetsvillig, mer social eller fått sin självtillit förstärkt?

- Är då kunskapsmål artskilda från andra mål? (DK)

- Det går inte att skilja dessa mål från varandra i den praktiska undervisningen, menade GS.

- Hur ska då utbildningen relateras till samhället, frågade RH. GS säger att vi vill anpassa elever till samhället och samtidigt skapa innovatörer. Det kan vi inte göra! Läroplansskribenterna har ju en viss politisk inställning. Du kommer alltså här in på vilken samhällsvärdering du har. Så länge vi har ett system som det svenska med många politiska uppfattningar, kan man inte precisera målen. Man vill inte precisera målen.

- Det blir också ohyggligt politiskt kontroversiellt om vi vill precisera vad vi menar med kunskap - att ange mål för kunskapen. Avslutade Daniel Kallós.

GRUPP B:

R_1	Aldemark, Lars
	Elison, Åke
	Graham, Peter
	Höghielm, Robert
	Kallós, Daniel
O	Levin, Lennart
	Silverberg, Bengt
	Solstad, Arne
	Svanberg, Gunilla
R_2	Svingby, Gunilla
	Ölund, Ulf

ETHICS AND EPISTEMOLOGY

Group B
(Lennart Levin)

Is there a sufficient relationship between what we know about teaching and what we know about learning to justify gathering student information (including test performance data) in order to correct or improve instruction?

It is reasonable to believe that borrowing from the legal profession, e.g., rules of evidence, procedures of adversarial crossexamination, might be useful in the evaluation of educational programs?

How important is it to identify the values held by the evaluator? Should safeguards against bias, narrowness, or irrelevance be built into the designs?

SAMMANDRAG AV DISKUSSIONERNA I GRUPP B (73.10.23)

Gunilla Svingby

Som startpunkt för diskussionen definierades epistemologi. Om "ontomologi" gäller varats - kunskapens - inre osäkerhet, gäller "epistemologi" den osäkerhet som finns mellan kunskapen och kunskapssökaren.

Diskussionen rörde sig huvudsakligen kring den första av de tre frågor som formulerats i anslutning till ämnet: ...
Centrala frågor blev då av typen: Vilken vetenskapsteoretisk syn utgår Bob Stake från när han diskuterar utvärdering? Hur ser vi själva på begreppet kunskap? Hur uppfattar vi verkligheten, vilka föreställningar har vi gjort oss om denna?

Den allmänna uppfattningen var att Stakes bidrag i grunden representerar ett logiskt-positivistiskt vetenskapsideal. Evaluerarens uppgift uppfattas vara att lyfta fram de skilda delar av "sanningen" som de olika deltagarna i ett undervisningssammanhang (program) representerar, för att därefter lägga samman biterna till något så fullständigt som möjligt, till något som kommer så nära "sanningen" som möjligt.

Det specifika med Stakes insats är att han inte accepterar att evalueraren ensam (med hjälp av test etc.) kan åstadkomma den sanna bilden av verkligheten. Han accepterar inte evaluerarens bild som den enda bilden av verkligheten utan söker även nå deltagarnas versioner. Det finns likheter här med den hermeneutiska ansatsen, men där denna låter deltagarna själva formulera mål och planera och genomföra kontrollen av den, har Stake uppdragsgivarens mål för evalueringen och en evalueringsplan.

Betoningen av "sociala kvaliteter" hos evalueraren och kravet på insikt i det program man evaluerar såg man som ett inslag som kunde vara uttryck för en annan vetenskapssyn än den positivistiska.

I den följande diskussionen om deltagarnas ståndpunkter ställdes mot varandra två uppfattningar. Å ena sidan hävdades att evaluering kan göras relativt fri från evaluerarens värderingar - "man kan ju göra objektiva prov". Det är möjligt att ge en "objektiv bild av verkligheten".

Företrädare för en annan grundsyn underströk det omöjliga i att studera fenomen (undervisning) som i grunden är yttringar av människors värderingar, utan att lägga på sina egna värderingar. Lärarens, läroboksförfattarens, evaluerarens uppfattning om vad kunskap är, hans syn på samhället och människan kan inte undgå att färga av sig på den undervisning han bedriver eller de provfrågor han konstruerar.

Beträffande relationen inlärning-undervisning framhålls från en utgångspunkt, att det finns lagbundna samband mellan "aktion och reaktion", mellan undervisningens organisation och elevernas inlärning. Den pedagogiska forskningens - och evalueringens - uppgift är följaktligen att söka "någorlunda styrkta samband" som kan generaliseras till nya sammanhang.

Utifrån en annan grundsyn framhölls att pedagogisk forskning inte studerar naturlagar eller fenomen i sig utan artificiella problem, problem som är skapade av pedagogerna själva, vilket bl.a. medför att vi inte kan förutsäga undervisningens resultat utifrån inlärningsteorier eller utifrån kännedom om inlärningslagar.

Man tvekade att erkänna något samband mellan inlärning och undervisning utan föredrog att anlägga ett idiosynkratiskt betraktelsesätt - varje situation är unik, det är fråga om individuella fenomen, som inte upprepas. Då det "efter 70 års pedagogisk forskning inte går att presentera en enda pedagogisk sanning" är det dags att söka en annan utgångspunkt för denna forskning. Behovet av generaliserbara resultat ifrågasattes. Man menade att på det lokala planet finns föga behov av generella anvisningar. Där skulle det vara bäst om lärare och elever i samråd bestämde vad och på vilket sätt de vill ha sin undervisning ordnad. Pedagogisk forskning med syfte att ge generellt användbara råd skulle egentligen vara obefogad - evalueringens och därmed också forskningens angelägenhet är att hjälpa den enskilda skolan att få sin situation att fungera.

Från ytterligare ett annat håll hävdades, att det är omöjligt att betrakta relationen inlärning - undervisning på ett så definitivt sätt. För att kunna resonera och ta ställning till grad och art av samband mellan dessa krävs att man söker definiera begreppen. Vad skiljer undervisning från inlärning? Man kan betrakta undervisning som ett kontiniuum

av allt mer omsorgsfullt planerade och arrangerade situationer för att möjliggöra inlärning. De enskilda unika situationerna kan utnyttjas för att bygga upp generaliserbar kunskap som kan användas vid planerandet av nya situationer.

Den intensiva diskussionen gav belägg för uppfattningen och visade att grundsynen har en avgörande betydelse för den roll man vill ge pedagogen-evalueraren och för det sätt man vill ställa och angripa problem.

GRUPP C:

Askling, Berit
Björsjö, Märta
Drambo, Leif
Gustafsson, Jan-Eric
Johansson, Martin
Kärrby, Gunni
Lepik, Andrus
Lindberg, Maj-Britt
R_2 Nilsson, Lennart H S
O Nyström, Astrid
R_1 Wenestam, Claes G

DEVELOPMENT OF INSTRUCTION

Group C
(Astrid Nyström)

Group D
(Kurt Gestrelius)

For any particular evaluation study, what are the bases for deciding how much emphasis to place on getting explicit statements of program objectives?

What special evaluation procedures are needed, if any, for formative evaluation?

Would evaluation studies be improved if there were a more explicit specification of the priorities to be placed on different instructional objectives?

Is a portrayal anything more than an entertaining and elaborate description of the program? What definitions or rules should we have for development or use of portrayals?

SAMMANDRAG AV DISKUSSIONERNA I GRUPP C (73.10.22)

Claes G. Wenestam

Vad nytt i Stakes modell?

Astrid Nyström: Stakes modell påminner om 50-talets resonemang.

Andrus Lepik: Det upplevs som nytt att någon vågar säga ifrån. Hittills har man utgått ifrån statistik.

Jan Erik Gustafsson: Man har gått ifrån de gamla resonemangen och flyttar upp en nivå till "nästa vetenskap".

A.N.: "Nästa vetenskap"?

Leif Drambo: Är det vetenskapligt, implicit kommer utfallen av evaluatorns värderingar. Det är nytt.

A.L.: Vad är vetenskap? Ur meriteringssyfte kan det vara dubiöst.

A.N.: Kan man komma idag i Sverige med sådana rapporter som vetenskapligt.

A.L.: Som pedagogiskt utvecklingsarbete - ja, men som pedagogisk forskning - nej.

Martin Johansson: Det är den status som Stake har, världsberömd etc. som gör att han kan komma med sådana här resonemang. Först måste man dokumentera sin förmåga enligt vedertagna vetenskapliga idéer. När man väl har blivit accepterad kan man komma med andra "lösare" resonemang.

Lennart Nilsson: På grund av de existentiella förutsättningarna, man måste ha mat etc., i den accepterade vetenskapliga miljön, präglas vad som kommer fram i rapporter och i designer etc. och vad som meriterar. I princip är det väl inget nytt synsätt. Se t.ex. på de antropologiska utvecklingsarbeten som finns. Det finns redan sådana här resonemang. Ett nytt klimat har kommit. Nytt tänkande, processinriktat i skolan m.m. är det som har möjliggjort det här. Möjligheten beror på vad som beslutsfattaren anser vara viktigt. Detta möjliggör att sådana här resonemang kommer fram.

A.N.: Tycker du att det är bra att Stake har kommit hit och för fram dylikt resonemang? Man ska inte hänga upp det på person o.dyl.

A.L.: Det följer på ett nytt synsätt, man överger t.ex. behavioristiska resonemang.

J-E. G.: Är det kunskap man söker med denna ansats den man söker i forskningen? Man skall inte sätta likhetstecken mellan forskning och evaluering. Det finns beröringspunkter, men ...

A.N.: Jag vänder mig till Märta Björsjö. Var tycker du den här typen av evaluering kommer in?

Märta Björsjö: Jag ser många tillfällen då den kan vara användbar. Varför kommer denna typ av evaluering just nu? Har man börjat tvivla på tidigare forskning?

J-E. G.: Kan man inte se det som en förändring i "Zeitgeist". En förändring på ett metaplan.

A.N.: Ett slags jämviktssträvan? En förändring i tidsandan?

M.B.: Knappast i tidsandan.

J-E. G.: Förändringar sker på många andra områden.

L.N.: Det berör de organisatoriska formerna. Forskningens krav på att kunna predicera. Om man tvivlar på den forskningsdesign, som använts, så går man över till en annan modell. Om de är nyttiga eller inte är omöjligt att säga på förhand. Man måste pröva dem först. Det finns en utveckling i evaluering, hos folk som sysslar med evaluering. Det är en trovärdighetskris i skolan som är förskräcklig.

A.L.: Huvudsyftet med evalueringar är att ge information till beslutsfattarna. Den behavioristiska approachen räcker inte till. Man är tvungen att gå över gränsen och studera processer och kanske tvingas att använda mjukdata. De bägge evalueringsmodellerna är ett komplement till varandra.

M.J.: Det skär sig mellan Stake och Pedagogiska institutionen här i Göteborg, som är högt teoretiskt inriktad, medan Stake vill ha mjukdatametoder med låg generaliserbarhetsgrad. Att bygga upp en teori gör man inte med en gång utan det är fråga om grundforskning under lång tid.

L.D.: Finns det något nytt i det här eller är det gammal skåpmat?

Gunni Kärrby: Hittills har vi valt våra problem utifrån våra värderingar etc. Vi får nu upp ögonen för andra företeelser, som andra människor/elever, lärare finner viktiga och som inte behöver sammanfalla med våra egna.
Trovärdighetskrisen. Jag upplever att resultat, som vi har forskat fram möts av ett - "jaså" - ute i skolorna. Det är värdefullt att få veta vad som anses viktigt på fältet. Det upplever jag som värdefullt och nytt.

A.N.: UPLM-gruppen som Andrus Lepik ingår i är ett konkret exempel på att vi försöker minska på gapet mellan olika grupper.

J-E. G.: Jag är böjd att inta samma ståndpunkt som Andrus Lepik att det är två ansatser. Det leder till en kognitiv konflikt. Man löser den genom att säga att evaluering är konkreta resultat för konkreta människor etc. men man kan inte överbrygga skillnaden så enkelt. Man har användning för denna kunskap även i den teoretiskt orienterade forskningen. Stakes datainsamlingsmodell och teoretiskt inriktad forskning kompletterar varandra.

A.N.: Är Stakes modell "subordinate" till teoretisk forskning?

J-E.G.: Nej, möjligtvis parallell. Det är beslutsunderlaget, som är det viktiga.

L.N.: Beslutsunderlag för vad? På systemnivå behövs en viss form av information, institutionsnivå behöver andra sorters data. Gruppnivå, individnivå behöver ännu annorlunda information. För olika typer av beslut är man i behov av olika typer av sätt att samla in data. När jag sa att det inte var något nytt i Stakes resonemang, menar jag, att det är nytt på så sätt att en person med Stakes ryktbarhet kommer hit och säger detta.

Berit Askiling: Jag tänkte ta upp trovärdighetskrisen. Jag upplever att sedan man varit i det här huset och sedan kommer ut som lärarutbildare blir det som att byta sida. Mycket av den traditionella forskningen sker utifrån forskarens speciella intressen. Hur länge skall en forskare tränas i traditionell forskning? Alla som forskar skall ju vara inom pedagogiska institutioner. Forskaren som person måste få framträda mer. Detta skulle minska trovärdighetskrisen.

Maj-Britt Lindberg: Jag tänker på Raoul Andrés rapport angående livet i en studentkorridor. Finns det några likheter mellan hans och Stakes resonemang?

J-E.G.: Det finns metodmässiga likheter men Raoul evaluerar inte.

A.L.: Vi inom forskarutbildningen skulle säkert vilja arbeta enligt Berit Asklings resonemang men det är svårt att gå utanför det etablerade systemet.

A.N.: Jag försöker provocera Märta Björsjö. Skulle du kunna tänka dig denna "approach"?

M.B.: Ja, det skulle kunna gälla lärare m.m. Det är viktigt för oss att följa det som sker, det nya. Det är just detta fria nya intervjuande, frågande, som jag anser vara värderfullt. Denna information som uppkommer, vilken är beskrivande, effektåterförande är mycket värdefull för en kontinuerlig utveckling.

A.N.: Jag tyckte att TWIN-CITY rapporten var för kort. Det var svårt att utgå från den som beslutsfattare, med dessa mjukdata som underlag. Jag skulle vilja ha mer hjälp av en rapport.

A.L.: Är det statistik, signifikansprövningar etc. som du vill ha eller är det mer bestämt tyckande i rapporten?

A.N.: Det måste vara en viss begränsning men ...

M.B.: Jag föreställer mig att man skulle behöva ett system, vissa genomtänkta modeller, vissa variabler, som man letar efter när man går ut. Inte minst för att kunna begränsa sig.

L.N.: Vad förutsätter Stakes modell? Ja, ett väldigt decentraliserat system. Hur skulle man kunna tillämpa dessa på SÖ? Han vill ha större frihet till människorna. Vilka beslut skulle människor på olika nivåer kunna få lov att besluta? Det är ett "risk-taking", som minskar ju högre upp man kommer i hierarkin. Det gäller att ha ett selektionssystem på varje nivå. Varje beslutsunderlag måste vara relaterat till vilka beslut som skall fattas. Uppe på SÖ:s nivå borde det inte förekomma någon information. Denna information skulle undan för undan ha sållats bort för annars försvinner SÖ i data.

M.J.: Men data som används som beslutsunderlag på hög nivå måste vara relaterade i hög grad med vad som händer på lägre nivåer. Relationen måste vara fastställd. Därför behövs pedagogiska teorier.

L.N.: Därför bör man inte fatta beslut på högre nivåer utan i stället delegera dem nedåt.

G.K.: Är det inte så att avrapportering med olika statistiska data gäller data som har hänt för 3 år sedan. Vad är det för värdefullt med sådan information? Är det också inte så att "responsive evaluation" är en annan slags forskning och vetenskap än den traditio-

nella och den bygger på en annan filosofi. När man säger att resonemanget har funnits tidigare kan det vara sant att det har funnits hos var och en men det nya är att det har bildats en begreppsapparat omkring resonemanget. Den traditionella forskningen är konserverande och befäster det vi redan vet, medan det nya resonemanget tillåter förändringar. Data blir inte 3-5 år gamla t.ex.

B.A.: Vi har på lärarhögskolan en s.k. Försöks- och Demonstrationsverksamhet. Om man hade ett allmänt resonerande, värderande etc. skulle man säkert ha en större benägenhet att använda andra forskningstraditioner än de vedertagna. Vi utbildar dem i statistiska metoder.

J-E. G.: Jag vänder mig till Astrid Nyström. Jag menar att vetenskaplighet är en ansats med vissa krav, t.ex. objektivitet, grad av reproducerbarhet m.m. Jag anser att utifrån dessa kriterier är skillnaden inte mellan evaluering och vetenskap, utan det är i användningen av data. I forskning strävar man att gå utanför situationen men det gör man inte i evaluering.

A.L.: Vilken tillämpning kan detta få? Inom UPLM-gruppen håller vi på att plocka fram en modell, som kan föras ut på lärarhögskolor. Där skulle vi sedan kunna använda denna mjukdatametod för att kunna förbättra den och få in värdefull information.

L.D.: Jag tror att man får tänka sig olika system på olika nivåer. Jag tror inte generaliserbarheten gäller om olika nivåer. Jag menar centralt, lokalt etc. På lägre nivåer behöver man informationsdjup, på högre nivåer behöver man i stället informationsbredd. Stakes modell ger informationsdjup. Den här ansatsen är värdefull på lägre nivåer. Framför allt därför att den sätter människan i centrum.

L.N.: När skall man börja vara tänkande personer i systemet? Jo, i förskolan. Men tyvärr finns det en utstötningsmekanism, som stöter ut kreativa personer. Man distribuerar kunskap och vissa metoder och utesluter andra.

M.J.: Jag tycker att rapporter även gör det svårt att fatta beslut på lägre nivåer. Stake förenklar resonemanget alltför mycket.

GRUPP C:

Askling, Berit
Björsjö, Märta
Drambo, Leif
Gustafsson, Jan-Eric
Johansson, Martin
Kärrby, Gunni
Lepik, Andrus
Lindberg, Maj-Britt
R_2 Nilsson, Lennart H S
O Nyström, Astrid
R_1 Wenestam, Claes G

DEVELOPMENT OF INSTRUCTION

Group C
(Astrid Nyström)

Group D
(Kurt Gestrelius)

For any particular evaluation study, what are the bases for deciding how much emphasis to place on getting explicit statements of program objectives?

What special evaluation procedures are needed, if any, for formative evaluation?

Would evaluation studies be improved if there were a more explicit specification of the priorities to be placed on different instructional objectives?

Is a portrayal anything more than an entertaining and elaborate description of the program? What definitions or rules should we have for development or use of portrayals?

SAMMANDRAG AV DISKUSSIONERNA I GRUPP C. (73.10.23)

Lennart H.S. Nilsson

Gruppdiskussionen kom att huvudsakligen centreras vid gruppmedlemmarnas reflexioner kring två av de fyra givna frågorna:

(1) For any particular evaluation study, what are the bases for deciding how much emphasis to place on getting explicit statements of program objectives?

(2) What special evaluation procedures are needed, if any, for formative evaluation?

Reflexioner kring ovannämnda frågor:

En av de avgörande frågorna angående ... the basis for deciding how much ... är varifrån man hämtar målen för evalueringen. Är det från s.k. "client needs" eller från de som bär beteckningen "audience"? Vad menas egentligen med "client"? Så vitt vi förstod betecknar "client" den (de) som anhåller om att få viss kurs, program o s v utvärderat, d v s beslutsfattare och resursfördelande personer. "Audience" däremot synes beteckna de, som genomgår kursen, programmet o s v, d v s "eleverna".

Bakom resonemanget om en "client" tycks ligga föreställningar, som är i parallellitet med resonemang inom vissa psykoterapeutiska skolor, om att "clienter" är i behov av hjälp för att klargöra "sina egna behov" ("needs"). Denna tankegång leder till tanken att evalueraren är i besittning av bättre förutsättningar än "client" att klargöra vad evalueringen skall syfta till, eller rättare sagt olika deleffekter med ett evalueringsarbete. Denna tankegång verkar mycket rimlig då evalueringen i huvudsak syftar till att förbättra undervisningsprocesser, d v s då evalueringen har ett formativt syfte (Se Bloom, Hastings, Madus 1971, s. 177. Handbook on formative and summative evaluation of student learning).

En väsentlig slutsats är således att programmets mål inte är identiskt med evalueringens mål, utan en del i evalueringens mål, och detta i synnerhet då evalueringen har ett stort uttalat formativt syfte.

Frågan angående var tyngdpunkten i ett evalueringsarbete skall ligga och i vilken utsträckning som den genom evaluering erhållna informationen skall användas formativt eller summativt kom att aktualisera frågan.

Vad är resultat?

I diskussioner kring detta problem kan man klart se att det föreligger en icke obetydlig skillnad i uppfattningen mellan å ena sidan personer, som i sitt vardagliga arbete är sysselsatta med beslutsfattande på central nivå och å andra sidan personer, som i forsknings- och/eller evalueringsarbete är upptagna med att finna strategier och tekniker för att kartlägga vilka omständigheter som försvårar eller underlättar genomförande av programarrangemang, d v s betingelser i och kring undervisningssituationer samt förändringar av dessa under undervisningens fortskridande.

I det förra fallet är "resultat" i allmänhet vetskapen om vad elever kunde prestera i form av vissa kunskaper och färdigheter samt ibland vissa sociala vanor och attityder i slutet av en utbildning. I det senare fallet betraktas vetskapen om programarrangemangens faktiska utformning och upplevda konsekvenser såsom relevanta resultat.

Man finner således om man hävdar att evalueringens mål inte är identiskt med vissa indikatorer på beteendeförändringar i relation till programmets mål ett klart behov av att klargöra "clientens needs".

Det ovan refererade förhållandet aktualiserade två problem vilka ventilerades inom gruppen. Det ena sammanhängande med evaluerarens ställning gentemot uppdragsgivare, det andra berörde evaluerarens kunskaper angående evaluering.

Följande huvudpunkter ventilerades beträffande evaluerarens ställning gentemot uppdragsgivare:

(1) Den centrala frågan för oss som arbetar inom det svenska utbildningssystemet är huruvida den "client-centrerade" "uppdragsgivarmodellen" är tillämpbar. Med andra ord om analogin mellan "client", som hjälpes att finna sina "needs" är adekvat då beslutsfattare inte består av en person som agerar utifrån sina implicita eller explicitgjorda behov utan är representanter för olika intresseinriktningar och då beslutsfattares ställningstagande är konsekvenser av i nuet existerande politiska styrkeförhållanden. Detta förhållande blir särskilt observerbart då evalueringsarbetet är nära anknutet till planerade och/eller nyligen genomförda utbildningsreformer.

(2) En väsentlig fråga är vilken aktiv roll evalueraren får och kan spela då det gäller att klargöra för "client" relevanta "needs".
Vilken roll han/hon får spela är till viss del beroende på vilka kunskaper som vederbörande är i besittning av.
I detta sammanhang kan naturligtvis utbildningen av evaluerare närmare diskuteras. Det sammanfattande resultatet av denna diskussion synes vara möjlig att sammanfatta på följande sätt:

Den blivande evalueraren bör inom ramen för forskarutbildningen erhålla en fördjupad kunskap om och träning i att utföra olika "typer" av evaluering. Det gäller dels generella strategier, dels specifika tekniker inte olika evalueringsinriktningar. En allmän enighet förelåg inom gruppen om att man inom "högre utbildning" tenderar att överdimensionera träning i vissa evalueringstekniker och av "outredda anledningar" försummar andra. Ett problem, som tycks sammanhänga med den naturvetenskapliga agrikulturella forskningsdesignens dominerande ställning inom beteendevetenskapliga discipliner.

Det ovan anförda ger uttryck för gruppens tro på möjligheterna att evaluerare i besittning av utökad kunskap om evaluering skulle kunna få större möjligheter att med "clienten" diskutera vad som bör ses som relevanta resultat av alternativa evauleringsarbeten och sedan få ansvara för val av strategi och tekniker, som utifrån evalueringens syften bäst passar i det aktuella fallet gentemot såväl "client" som "audience".

I realiteten föreligger en koppling mellan (1) och (2) eftersom det i realiteten är samma beslutsfattare som beställer evaluering och som har möjligheter att prioritera resurser till evaluerarutbildning.

GRUPP D:

Andersson, Lillemor
André, Raoul
R_2 Colliander, Marianne
Elliström, Per-Erik
O Gestrelius, Kurt
Hendel, Gunnar
Jonsson, Maud
Köhler, Eva Marie
Marton, Ference
Olsson, Ingela
R_1 Rosén, Kerstin
Stencrantz, Annika

DEVELOPMENT OF INSTRUCTION

Group C
(Astrid Nyström)

Group D
(Kurt Gestrelius)

For any particular evaluation study, what are the bases for deciding how much emphasis to place on getting explicit statements of program objectives?

What special evaluation procedures are needed, if any, for formative evaluation?

Would evaluation studies be improved if there were a more explicit specification of the priorities to be placed on different instructional objectives?

Is a portrayal anything more than an entertaining and elaborate description of the program? What definitions or rules should we have for development or use of portrayals?

SAMMANDRAG AV DISKUSSIONERNA I GRUPP D. (73.10.22-23).

Marianne Colliander och Kerstin Rosén

Inledning

Den diskussion som fördes i gruppen under de två dagarna kom att ganska starkt påverkas av den osäkerhet om Stakes avsikter och teoretiska förankring, som vi upplevde. Detta bidrog till att diskussionen blev ganska osystematisk och splittrad, och vissa teman återkom vid flera tillfällen. Trots den tid vi ägnade åt att försöka klargöra Stakes grunduppfattning lyckades vi inte nå fram till någon gemensam bild av denna. Redogörelsen kommer att beröra även de diskussioner vi hade runt Stakes kunskaps- och vetenskapsteoretiska synsätt, något som till en del låg utanför de frågeställningar som egentligen var avsedda att behandlas. För att underlätta för läsaren har vi försökt att något redigera materialet och presentera den diskussion vi förde på ett något mer systematiskt sätt.

Gruppens arbete kom under första dagen att huvudsakligen inriktas på tre givna frågeställningar, nämligen 1) Vad är nytt i Stakes approach, 2) Vad finns det för skäl till att den av Stake representerade typen av evaluering inträffar just nu samt 3) Vad ser vi för tillämpningsmöjligheter för responsiva evaluation.

Vad är nytt?

Diskussionen startade med att vi försökte se om det fanns några skillnader mellan illuminative och responsive evaluation. Vi kom därvid gemensamt fram till att vi såg de två begreppen som liktydiga.

Efter att sålunda ha undanröjt ev. definitionsmässiga svårigheter gick diskussionen över på frågan om vad som är nytt i Stakes betraktelsesätt. Samtliga gruppmedlemmar var överens om att Stakes idéer i själva verket inte var särskilt nya eller revolutionerande. Bl a diskuterades Trankell, som konstaterades ha kommit ett steg längre än Stake. Antropologerna såg vi som ett annat exempel på forskare som arbetat med "soft data".

För att se om vi kunde finna några nya aspekter i den responsiva evalueringen koncentrerade vi oss därefter på en mer penetrerande analys av Stakes idéer. Vi kom därvid fram till att det framför allt rörde sig om en tyngdpunktsförskjutning i jämförelse med traditionell evaluering, där responsive evaluation har koncentrationen lagd på en intuitiv beskrivning, som syftar till att ge mer direkt kunskap genom eget upplevande och egna erfarenheter. Härav följer att Stake argumenterar för ett beskrivande presentationssätt, som syftar till att ge en helhetsbild av processen och inlärningsmiljön med utnyttjande av mer onkonventionella metoder, hämtade t ex från skönlitteratur.

På den positiva sidan såg vi ett visst nyhetsvärde i den betoning av öppenhet och förutsättningslöshet i sättet att närma sig problemet, som bör kännetecknа forskaren. Detta kan ge upphov till att nya faktorer observeras och nya variabler och registreringsmetoder, som växer fram under processens gång, uppmärksammas.

Vi ser det också som något delvis nytt att den responsiva evalueringen är mer avpassad för en beslutsfattande avnämargrupp än traditionella evalueringsmetoder, som ofta riktar sig till ett forskarauditorium.

Stakes betoning av teaching och evaluation som "an act" snarare än som en vetenskaplig studerbar kompetens har vissa, som vi såg det, negativa konsekvenser. Bl a blir han starkt bunden till individnivån, vilket försvårar möjligheterna att generalisera till andra situationer. Hans åsidosättande av efterhandssystematiseringen försvårar möjligheterna att dra enhetliga explicita slutsatser och lämnar åt var och en att bilda sig en egen uppfattning. Detta medför att han får svårt att försvara de slutsatser han drar. Vi efterlyste här evidens.

Vi diskuterade vidare behovet av att ha ett speciellt syfte i evalueringen för att kunna avgöra vilken typ av data - kvantitativa eller kvalitativa - som är nödvändiga för att kunna dra slutsatser. Även här fann vi en svaghet i Stakes modell. Problemet med representativitet var en annan risk, som vi upplevde kunna vara av kritisk betydelse vid användandet av denna modell.

Varför nu?

Diskussionen av den tidsmässiga aspekten av Stakes modell blev ganska kort. Vi såg det hela som ett typiskt uttryck för den tes - antites process, som är karakteristisk för all idéutveckling.

Den omsvängning i vetenskapsteorin mot ett mer holistiskt betraktelsesätt, som f n kan förmärkas, de social-politiska idéerna i vår omgivande miljö, tidsandan såg vi som bidragande orsaker. Påverkan från andra discipliner, som t ex antropologin, var en annan orsak. Ytterligare en påverkande faktor såg vi komma från det konkreta arbetet med evaluering. Genom systematisering av successivt förvärvade kunskaper och erfarenheter når man så småningom fram till insikt om att de data man hittills arbetat med inte är tillräckliga och tillfredsställande i förklaringssyfte, och man söker sig över till andra typer av data.

Vi efterlyste en integrering av traditionell och responsiv evaluering för att i praktiska situationer nå bättre täckning av problemområdet (soft plus hard data).

En stunds diskussion ägnades åt frågan huruvida tes - antites utvecklingen ledde framåt mot nya kunskaper eller endast var en sorts springande i ekorrhjulet utan förmåga att generera ny kunskap.

Tillämpningsmöjligheter

Denna fråga kom p g a tidsbrist att behandlas mycket ytligt. Framför allt såg vi responsive evaluering ha värde i befattningsutbildning och som komplement till traditionell evaluering. Företrädesvis såg vi användingsområdena knutna till formativa evalueringssituationer.

Diskussionen kom sedan att glida in på ett tidigare berört område, nämligen möjligheten att över huvud taget kunna komma helt förutsättningslös till en evalueringssituation. Vi konstaterade härvid, att det för att man skall kunna arbeta på ett sätt som av andra upplevs som förutsättningslöst egentligen krävs att man är extra väl förberedd.

Diskussionsämnet för andra dagens grupparbete var "Development of instruction".

Första delen av tiden ägnade vi åt en diskussion av vilka typer av utbildningsmål vi har och hur vi definierar dem. Vi konstaterade att man genom en behovsanalys kunde särskilja två typer av mål - nämligen utbildnings/kunskapsmål (instructional objectives) och processmål (intended transactions). Vi såg också en tredje målkategori - systemmål - vilka hänför sig till ekonomiska och administrativa aspekter av utbildningen. Den hittillsvarande forskningen har framför allt koncentrerat sig på de enklare, mätbara kunskapsmålen, medan svårigheterna att hitta användbara mått på processmålen bidragit till att de lämnats ganska mycket därhän.

Som vi såg det prioriterar Stake processevaluering men utan att ha några förutbestämda, explicita processmål. I den diskussion vi förde runt detta var vi ganska eniga om att vi inte såg det möjligt att genomföra en evaluering utan någon form av mål som styrinstrument. Vi trodde inte heller att Stakes avsikt var att förespråka en "mållös" evaluering. Såväl processmål som kunskapsmål har berättigande. Om man väljer att utvärdera det ena eller det andra eller en kombination av dessa måste bli beroende av syftet med evalueringen. Finns explicita kunskapsmål i läroplanen skall dessa givetvis inte ignoreras.

Svårigheten, för att inte säga omöjligheten, att fullständigt beskriva en viss del av verkligheten bidrar till att man bör prioritera vissa aktiviteter, som är av kritisk betydelse för måluppfyllandet.

En viss form av kategorisering och systematisering av data bedömde vi vara nödvändig även vid användning av den stakeska portrayal modellen. Detta, menade vi, beror dels på den ovannämnda svårigheten att täcka av den totala situationen, dels på behovet att utnyttja de kunskaper, erfarenheter och värderingar vi har från tidigare situationer.

Vi kom så in på en diskussion om Stakes vetenskapsteoretiska förankring. En av gruppmedlemmarna skisserade följande figur, vilken övriga gruppmedlemmar tog som sin:

TYPE	AIM	GENERALIZABILITY
Analytic evaluation	Explain	1) in the specific case
(conclusion oriented)	WHY	2) with high degree of generalisability
Nonanalytic eval.	Describe	3) for the specific case
(decision oriented)	WHAT	

2) har varit den traditionella forskningens syfte. Den medger inte förklaringar på individnivå, vilket är en svaghet. Stake vill vi placera in under 1) och 3). Hans styrka är att han försöker beskriva orsaker för varje enskilt fall. Samtidigt förlorar han emellertid möjligheten att överföra kunskap från en situation till en annan likartad situation.

Några av gruppmedlemmarna ville betrakta Stake som naiv empirist. Alla var överens om, att det var en svaghet att han inte ville (kunde?) ge uttryck för explicita vetenskapsteoretiska värderingar.

Vid en diskussion om huruvida portrayals är något mer än underhållande deskription fann vi oss kunna konstatera, att avnämargruppernas behov här var det kritiska kriteriet. Såväl mängden som strukturering av data bör styras av dessa behov. Om syftet är att försöka nå ut till en stor, icke-fackkunnig publik, kan portrayals sannolikt ge en stimulerande och förståelig läsning.

Ett övervägande om att ev presentera materialet på mer än ett sätt kan vara lämpligt i situationer, där man önskar nå mer än en avnämargrupp. Det under måndagens diskussion ställda kravet på evidens kan emellertid inte nog poängteras i en rapport som riktar sig till sakkunniga beslutsfattare.

Som avslutning ställde vi oss frågan om evalueringsteori kanske i själva verket är en teori om beslutsfattare eller beslutsfattande. Frågan lämnades öppen och överlåtes med varm hand till läsarna att besvara.

GRUPP E:

Andersson, Bengt Erik
R_2 Gustafsson, Christina
R_1 Johansson, Gunnel
Levén, Sören
Lindblad, Sverker
O Magne, Olof
Petersson, Lennart
Rolf, Magnus
Rovio-Johansson, Airi
Svensson, Harry

RESEARCH METHODS

Group A (Esse Lövgren) — Should the evaluator be reluctant to use unsolicited observations and casual opinions as data? Can we validate or corroborate such data?

Group E (Olof Magne) — Should the methods of evaluation differ for different subject-matters, for different agegroups, for different types of students and teachers?

Group F (Karl Gustaf Stukát) — Will classroom interaction analyses, e.g., Flanders, be useful in evaluation studies?

Group G (Lars Gunnar Holmström) — What do ideas about responsive evaluation suggest for the training or retraining of educational researchers?

SAMMANDRAG AV DISKUSSIONERNA I GRUPP E. (73.10.22)

Gunnel Johansson

Gruppens diskussion startade med att närma sig själva evalueringsbegreppet, genom att klargöra dess historia och betydelse, och frågor som "varför har vi en evaluering" och "till vad skall den tjäna" försökte besvaras.

Ur diskussionen som följde framkom att syftet med en evaluering är att förbättra och att söka problem men den utgör även en bas för information. Det framkom även att själva evalueringsbegreppet utgör en komplex natur av företeelser som samspelar med varandra, och själva evalueringen är en typ av värdering i förhållande till uppsatta normer.

Därefter följde ett försök till en betydelseangivelse av begreppen formativ och summativ evaluering. Ur detta försök framkom att den summativa evalueringen endast är en liten tidssektor och problemet huruvida det är en summativ evaluering vid tidpunkt t_2 likaväl som vid tidpunkt t_1 ledde till slutsatsen att den summativa evalueringen är en abstraktion. Slutligen enades gruppen om att både formativ och summativ evaluering innefattas i Stakes responsive evaluation.

Gruppen övergick därefter till att dela upp evalueringen i två metoder som benämndes den traditionella metoden och responsive metoden och kom på detta sätt in på den första frågeställningen: "Vad är det nya i Stakes angreppssätt"?

Gruppen tyckte att den traditionella metoden är inriktad på en begränsad uppsättning mål, medan responsive metoden är uppmärksam på andra förändringar, m a o den har en helhetssyn (ett bredare perspektiv) och betraktar inte "händelser" som isolerade företeelser utan tar också hänsyn till vilka konsekvenser får det för t ex de övriga målen. Vad Stake främst har bidragit med är att han har ändrat på evalueringsproblematiken genom att det är problemet (en) som är det intressanta.

Nästa fråga som gruppen ställde sig var om Stake förkastar den traditionella metoden eller om responsive evaluering är superordinate. Därefter antog gruppen att Stake

menar att responsive evaluering är ett helt nytt alternativ.

Vari ligger det nya? Gruppen finner att det finns tre nya punkter:

1. att hela inlärningsmiljön är med (helhetssynen)
2. att processen finns med (program activities)
3. issues - upplevda problem (ökad medvetenhet, medbestämmanderätt)

Diskussionen handlade sedan om evaluering kontra forskning och Stakes uppfattning[1] är att evaluering är inte forskning, utan evaluering är ett agerande i miljön, och ett sätt att se och upptäcka det specifika.

Den slutsats som gruppen så småningom kom fram till angående den första frågeställningen var, att det är inte intressant att tala om traditionell metod kontra Stakes metod utan det är snarare en fråga om en metod att välja olika tekniker.

Den andra frågeställningen: Varför förekommer dessa "strömningar" just nu (1973)? eminerade i en diskussion om vårt vetenskapsideal och från diskussionen framkom att det är vårt strikt positivistiska vetenskapsideal som har hindrat oss från att använda vissa instrument. Vidare framkom att ett annat sätt (perspektiv) är att se hela systemet (ex skolan) som ett problem m a o att ta med hela omgivningen.

Därefter kom gruppen in på vad man brukar kalla de tre F-en, nämligen förklaring, förståelse och förändring. Gruppen ansåg att Stake har börjat röra sig från förklaring mot förståelse, men den stora frågan är hur åstadkommer vi en förändring. Detta berör ej Stake, vilket av gruppen upplevdes som en brist.

En "modell" skissades därefter för att försöka förklara "förskjutningen" och den konstituerades av vårt vetenskapsideal, våra problem (intentioner), våra instrument och vår kunskap (i snäv mening t ex s - r) och p g a syftet med modellen lades i detta fall evalueringen i mitten. Modellens nyss uppräknade komponenter ansågs påverka varandra och i sin tur evalueringen i en ständig växelverkan. Den första "förändring" som

---

[1] enligt gruppen

kunde iakttas är att vi nu skiljer på kunskap i snäv och vid mening, d v s vi lämnar alltmer de traditionella kunskapsidealen. Vidare har vi blivit mer processinriktade och ser individen som en helhet. Detta måste i sin tur leda till frågor som "vad vill vi med skolan"? och "vad vill andra med skolan"? Gruppen tyckte att vad som saknas hos Stake är en teori om hur skolan fungerar och hur teorin knyts till evalueringsprocessen och här slutade gruppens diskussion angående denna fråga.

Den sista frågeställningen stimulerade till ett försök att applicera responsive evaluering på ett "konkret" fall och upplevdes av en del gruppmedlemmar som starkt avhängig av vilken uppfattning (teori) man har om hur skolan skall fungera. Andra medlemmar i gruppen ifrågasatte däremot om man kan acceptera "allt" som evaluatören finner väsentligt att evaluera. Frågan om vem som skall upptäcka problemen upplevdes som väsentlig men berördes endast flyktigt.

Samskapsskolorna i Malmö togs upp för där sker utvärderingen enligt en indikatorsmodell och på detta exempel gjordes ett försök att applicera responsive modellen. Problemet som diskussionen rörde var bl a vem som skall delges informationen. Därefter föreslogs skolans övergripande mål och betygsfrågan som intressanta frågor att relatera till responsive evaluering. Här var tiden för gruppdiskussionen emellertid ute så något egentligt försök att tillämpa responsive evaluering på ett exempel hann gruppen tyvärr inte med.

GRUPP E:

	Andersson, Bengt Erik
R_2	Gustafsson, Christina
R_1	Johansson, Gunnel
	Levén, Sören
	Lindblad, Sverker
O	Magne, Olof
	Petersson, Lennart
	Rolf, Magnus
	Rovio-Johansson, Airi
	Svensson, Harry

RESEARCH METHODS	Should the evaluator be reluctant to use unsolicited observations and casual opinions as data? Can we validate or corroborate such data?
Group A (Esse Lövgren)	
Group E (Olof Magne)	Should the methods of evaluation differ for different subject-matters, for different agegroups, for different types of students and teachers?
Group F (Karl Gustaf Stukát)	Will classroom interaction analyses, e.g., Flanders, be useful in evaluation studies?
Group G (Lars Gunnar Holmström)	What do ideas about responsive evaluation suggest for the training or retraining of educational researchers?

SAMMANDRAG AV DISKUSSIONERNA I GRUPP E (73.10.23)

Christina Gustafsson

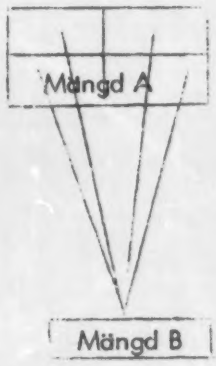
Genomgående präglades andra dagens livliga diskussion av meningsskiljaktligheter, vilka krävde förklaringar och motiveringar. Av denna anledning blev diskussionen stundtals ganska perifer med avseende på de givna frågeställningarna rörande metodfrågor vid evaluering.

Gruppen startade med en summering av föregående dags diskussioner. Fortfarande talade man om evaluering i termer av å ena sidan traditionell evalueringsmetod och å andra sidan responsive evaluation ("Stakes metod"), där den förra sågs som ett specialfall av den senare.

En kort diskussion om mål och målformuleringar vid utvärdering följde sedan. Bl a såg man en risk i att mål efter utvärdering relateras till specifika processmönster, eftersom detta kan medföra en styrning av processen mot symptom och ej mål. En exemplifiering borde i detta sammanhang vara tillräcklig.

En av de givna frågeställningarna, nämligen den som rörde olika eller identiska evalueringsmetoder för olika ämnen, olika åldersgrupper och olika elever och lärare (frågeställning 2), gav upphov till en längre diskussion. Generellt anslöt sig gruppen till ett inlägg, som tidigare redovisats under symposiet, nämligen att problemets karaktär skall vara avgörande för vilket förfarande som kommer ifråga vid evalueringen. En av medlemmarna i gruppen presenterade i anslutning till detta en principiell arbetsgång enligt följande:

1. Vi har en mängd händelser i verkligheten
2. Vi arbetar med en "modell", d v s vi har intentioner med evaluering.
3. Vi har inte möjlighet att i vår evaluering täcka hela fältet, varför vi väljer en strategi för evaluering.



4. Vi samplar ett antal händelser som vi avser studera och får en ny mängd.

5. Vi har använt indikatorsmetod.

Förslagsställaren till ovanstående arbetsgång betraktade denna tillämplig såväl vad gäller traditionell evalueringsmetod som den av Stake presenterade metoden. Trots att den senare av gruppen betraktades som en global metod ansågs den alltså fortfarande vara en indikatorsmetod.

Resonemanget enligt ovan väckte i gruppen reaktioner huvudsakligen på tre punkter:

- Varifrån kommer "modellen" i steg 2 och 3 och vad står den för? Just detta steg ansåg några gruppmedlemmar som det mest avgörande. Vilken "modell" som här blir aktuell sågs som en effekt av vilket perspektiv och vilket vetenskapsideal evaluatören anlägger på verkligheten. Det skulle innebära att det finns många former av "modeller" och varje sådan "modell" styr samplingstekniken, alltså vilka element som väljes ut och vilken relation som antas finnas mellan elementen. På så sätt får "modellen" konsekvenser för vad vi kan uträtta. Samplingstekniken medför också att vi missar strukturen.

- Är den presenterade arbetsgången tillämpbar på "Stakes metod"? Framför allt i stegen 2-3 såg man en skillnad. Å ena sidan förekom uppfattningen att Stake helt hoppade över dessa två led genom att förutsättningslöst bara gå ut och se vad som händer. Å andra sidan fanns de, som tyckte att led 2 och 3 fanns med i "Stakes modell", men med den skillnaden att deltagarna i stället för evaluatören väljer elementen.

- Större delen av gruppen vände sig vidare mot avsaknaden av en växelverkan mellan 2:a och 3:je steget. Förslagsställaren tillfogade förklaringen att hela arbetsgången kunde betrakas som en cykel, som i sin tur kunde följas av en eller flera cykler, även om verklig-

heten av budgetsmässiga skäl ofta endast tillät en cykel och ibland bara delar av en cykel. Gruppen såg det i samband med detta nödvändigt att evaluatören från början planerade åtminstone 3-4 cykler i syfte att tillgodose den väsentliga verklighetsanknytningen.

Sammanfattningsvis såg gruppen med anledning av den ursprungliga frågan varje evalueringsstrategi som unik och icke generaliserbar. Det viktigaste är att evaluatören kan bestämma, vilken som är den bästa möjliga metoden inom givna ramar.

Gruppen övergick sedan till att diskutera, huruvida den typ av data som Stake introducerar är möjliga att validera (frågeställning 1). Gruppdeltagarna såg det omöjligt att göra detta enligt traditionella metoder. "Stakes strategi" kräver enligt några gruppmedlemmars mening tolkningsmoment. I annat fall kan exempelvis inte en förståelseprocess hos elever valideras. De gruppmedlemmar som inte instämde i denna åsikt betraktade evaluering som omöjlig såvida det inte var frågan om till exempel en utsaga om förståelse, som frekvensmässigt kan valideras. Gruppen fann det i detta sammanhang nödvändigt att man börjar fråga sig, varför man evaluerar och redan från början beaktar generaliserbarheten och implementationsproblematiken.

Två frågeställningar (frågeställningarna 3 och 4) behandlades på grund av tidsbrist endast flyktigt i gruppen. Gruppmedlemmarna ansåg i enlighet med sin inställning att problemet bör vara bestämmande för evalueringsmetoden och att, om de frågor vi ställer kräver analys av klassrumsinteraktionen kan det vara befogat att använda en sådan typ av analyssystem. Den största faran gruppen såg i detta sammanhang var, att analyssystemet kunde bli styrande för problemet.

Frågeställning 4 avsåg nyare evalueringstrenders eventuella effekt på forskarutbildning och vidareutbildning. Gruppen ansåg det viktigt, att forskarutbildningen breddas och inte ses isolerad från vad som sker på lärarhögskolorna. Ett socialantropologiskt arbetssätt som innebär, att forskaren vistas i skolor och ökar kontakten med berörda parter borde enligt gruppens mening beredas plats. Gruppmedlemmarna trodde, att en sådan omvärdering av vad som nu göres inom givna ramar inte företrädesvis kräver en kunskapsförändring utan en attitydpåverkan.

Sammanfattningsvis förhöll sig gruppen mycket positiv till den evalueringsmodell Stake presenterat. Gruppmedlemmarna antog dock att tillämpningen av idéerna på grund av mångsidigheten kräver att medvetenheten hos lärarna ökas. Ty till sist är det läramas totala förhållningssätt, som är avgörande för hur utvärdering på denna nivå kan ske.

GRUPP F:

Berggren, Ingegerd
Björklund, Eskil
R_2 Dahlgren, Lars Owe
Ellmin, Roger
R_1 Ericson, Lars
Idman, Pekka
Jansson, Sven E B
Lundman, Lars
Nordén, Bengt
Sandgren, Björn
O Stukát, Karl Gustaf

RESEARCH METHODS	
Group A (Esse Lövgren)	Should the evaluator be reluctant to use unsolicited observations and casual opinions as data? Can we validate or corroborate such data?
Group E (Olof Magne)	Should the methods of evaluation differ for different subject-matters, for different agegroups, for different types of students and teachers?
Group F (Karl Gustaf Stukát)	Will classroom interaction analyses, e.g., Flanders, be useful in evaluation studies?
Group G (Lars Gunnar Holmström)	What do ideas about responsive evaluation suggest for the training or retraining of educational researchers?

SAMMANDRAG AV DISKUSSIONERNA I GRUPP F (73.10.22-23)

Lars Ericsson och Lars-Ove Dahlgren

Gruppen ägnade sig inledningsvis åt ett försök till begreppsbestämning med följande resultat:

Preordinated evaluation: "Traditionell" eller "förplanerad" evaluering, d v s en evaluering mot i förväg uppställda mål efter en i förväg uppställd plan.

Responsive evaluation: Vid "responsiv" evaluering finns möjligheten att "känna sig för". Metoderna för observation anpassas till programmets framåtskridande och justeringar i programmet kan ske direkt. Evalueringen sker följdaktligen inte mot i förväg uppställda mål utan fokuseringen ligger på själva processen.

Nedan redovisas i punktform ett försök till sammanfattning av de synpunkter som framfördes under gruppdiskussionen.

Vad är det nya med Responsiv Evaluering?

Rapporteringen görs avnämaranpassad. Man syftar till en mer levande helhetsbild.

Observation ges mycket hög prioritet i arbetet, ca 30 % av tiden, vilket är mycket mer än vid traditionell evaluering. Evalueraren tvingas tränga "verkligheten" in på livet.

Ger svar på andra frågor. En problemorientering möjliggöres. I stället för att fråga hur frågar man varför. Responsiv evaluering representerar ett annat vetenskapsideal med inriktning på förståelse.

En ny roll för evalueraren. Evalueraren måste kunna tränga in i stämningar och kunna återge dessa. Han/hon får genom att vi, i högre utsträckning, kommer att mäta socialt beteende ett annat ansvar. Han/hon kan inte som traditionellt välja variabler som innebär så lite ansvar som möjligt.

En förskjutning från den traditionella tyngdpunkten på målen. Eventuellt kan denna förskjutning innebära att de sociala målen (skolans övergripande mål) uppmärksammas i större utsträckning.

Objektivitetsproblemet. Forskaren saknar för närvarande kriterier för kontroll av sitt arbete enligt dessa "mjukare" metoder. Mot detta kan ställas att rapporteringsformen gör att subjektiva inslag kan upptäckas lättare än vid traditionell rapportering då subjektiviteten döljs bakom korrelationsmatriser etc.

Måste evaluerarен komma utifrån?

För (Evaluerarен kommer givetvis att lära känna situationen genom sina observationer och i viss mån bli deltagare varför termen "utifrån" inte är helt relevant)	Emot (Argumenten emot gäller evaluering på "mikro"-nivå)
Värdekonflikter upptäcks lättast om evaluerarен kommer utifrån. Läraren måste ses tillsammans med eleverna. Om läraren själv gör evalueringen gör han det utifrån ett eget perspektiv. (Systemnivå) Det vi evaluerar kan ha ett externt värde. När utbildningen tjänar som intagningsbiljett till olika privilegier måste någon form av "rikslikare" upprättas vilket motiverar att evaluerarен kommer utifrån.	Det är onaturligt att se evaluering som fristående från undervisning. Den som ska evaluera måste ha en viss grad av kunskap om det som ska evalueras och bör kunna göra något åt bristerna när de upptäcks.

Varför uppträder dessa strömningar nu och kan de pedagogiska institutionerna vara mottagliga för dem?

Det finns ett behov av att komplettera traditionella metoder. Man vill lämna eller utvidga de konventionella kunskapsidealen. De traditionella metoderna har använts länge och inte gett så mycket, vilket gör det nödvändigt att söka nya vägar. Det har bedrivits mycket improduktiv forskning.

Vi behöver innovationer för att överleva. Genom alternativa utvärderingsmodeller skulle man kunna upptäcka de människor som brukar kallas kreativa.

Den rena produktevalueringen blir för smal och går också emot demokratiska krav på medbestämmande för de som utsätts för programmen.

Vi är mottagliga därför att synsättet ger oss en möjlighet att ställa varför-frågor vilket utgör en grundförutsättning för teoribyggandet.

Metodiska frågor i samband med evaluering

Valet av evalueringsmetoder kan anses vara resultatet av avvägningar mot ett flertal aspekter. Bland sådana diskuterades främst:

1. Det är icke betydelselöst vem som utför evalueringen. I t ex USA bedrivs en stor del av verksamheten inom detta område av etablerade professionella forskare. I Sverige utförs en stor del av arbetet av forskarstuderande. Det är naturligt att traditionella vetenskapliga kriterier i hög grad styr genomförande och rapportering i ett fall av senare typ.

2. Det är av stor vikt för metodvalet vem uppdragsgivarna är. Beslutsfattare har i hög grad vant sig vid vissa undersökningstekniker genom en dubbelinriktad anpassningsprocess vis-a-vis forskarna. Allt för stora avsteg från denna tradition kan skapa misstänksamhet mot resultatens tillförlitlighet.

3. Den dominerande synen på evalueringens funktion och syften styr i hög grad metodvalet. Uppfattningar om undervisningens mål, planering och genomförande spelar härvid en stor roll för hur man väljer att betrakta evaluering. Avvägningar med avseende på fakta- eller färdighetsbetoning, betonande av sociala, affektiva eller kognitiva mål, utgör exempel på egenskaper hos undervisning som får konsekvenser för evalueringen.

Mot bakgrund av bl a ovanstående styrfaktorer diskuterades skillnader mellan de möjligheter som "preordinate" respektive "responsive" evaluering tillhandahåller.

Det var en allmänt utbredd uppfattning att det vore önskvärt att den "mjukare" och mer "avnämarvänliga" metodik och rapportering som Stake förknippar med responsive evaluering, kunde accepteras som vetenskapligt giltig. För åtminstone vissa syften, t. ex. explorerande studier, är det uppenbarligen olämpligt att använda den instrumentarsenal som traditionellt ingår i "preordinate" evaluering.

I samband med diskussionen om olika metodorienteringar återknöts till det tidigare resonemanget om "objektivitetsproblem" vid den uteblivna instrumentstyrning, som kan bli resultatet av förutsättningslös problemstyrning. "Objektiviteten" består i detta fall dels av representativitet, dels av noggrannhet.

Observationsschemata, test m.m. har framför allt sin styrka i noggrannheten. En förutsättning för en resulterande hög "objektivitet" är dock att man använder instrumenten på samma typ av fenomen som de utprövats för samt att dessa fenomen är väsentliga för den situationen som studeras. Vid användning av prefabricerade instrument löper man alltid risken att deformera verkligheten till överensstämmelse med instrumentets villkor.

I "responsive" evaluering ligger underförstått att hjälpmedlen skall växa fram ur den förutsättningslösa problemobservationen. Avsikten är m a o att metodiskt söka "skräddarsy" instrumenten. Detta hindrar dock inte att dessa instrument i en framtid kan komma att utnyttjas mera "konfektionsmässigt". Poängen i detta resonemang är att det efterhand blir omöjligt att metodiskt diskriminera "preordinate" och "responsive" evaluering eftersom en utbredning av "responsive" evaluering rimligtvis så småningom kommer att leda fram till en breddning och diversifiering av metodarsenalen vilket vore en högst lyckosam utveckling.

Forskning och evaluering

Gruppen upplevde påtagliga svårigheter att hålla isär tillämpningarna av begreppet "responsive" inom forskningsprocessen i allmänhet respektive evaluering i synnerhet. Robert Stake uppträdde förmodligen mer som evaluerare än som forskare vid detta tillfälle men samtalen gled vid åtskilliga tillfällen över mot en mer generell tillämpning

av begreppet "responsive". Eftersom pedagogisk forskning i icke liten utsträckning behandlar evaluering är det dock ofrånkomligt att egenskaperna hos evalueringsfilosofin "smittar av sig" på forskningsprocessens olika delar. En intressant fråga, som dock inte fick något svar vid diskussionerna, är hur sådana arbeten kommer att bedömmas såsom vetenskapliga verk.

Utanför forskningsdomänen i en snäv mening har man redan erfarenheter av ett "mjukare" arbetssätt. I Sverige gäller detta t.ex. gymnasieinspektörernas arbete och rapportering eller den s k diciplinnämndens funktionssätt.

GRUPP G:

R_2	Bladini, Ulla-Britt
	Carlsten, Yngve
O	Holmström, Lars Gunnar
	Lövgren, Horst
	Markesjö, Gunnar
R_1	Patriksson, Göran
	Rudemo, Lisbeth
	Schulenburg, Rosvita

RESEARCH METHODS	Should the evaluator be reluctant to use unsolicited observations and casual opinions as data? Can we validate or corroborate such data?
Group A (Esse Lövgren)	
Group E (Olof Magne)	Should the methods of evaluation differ for different subject-matters, for different agegroups, for different types of students and teachers?
Group F (Karl Gustaf Stukát)	Will classroom interaction analyses, e.g., Flanders, be useful in evaluation studies?
Group G (Lars Gunnar Holmström)	What do ideas about responsive evaluation suggest for the training or retraining of educational researchers?

SAMMANDRAG AV DISKUSSIONERNA I GRUPP G (73.10.22-23)

Göran Patriksson och Ulla-Britt Bladini

Efter en kort presentation av gruppmedlemmarnas bakgrund och verksamhetsområden startade diskussionen i anslutning till i huvudsak två av de tre föreslagna temata, nämligen vad som var nytt i Bob Stakes idéer om evaluering och varför dessa idéer började uppträda just nu. Eftersom dessa två frågeställningar var intimt kopplade till varandra var det svårt att hålla isär dessa problem under diskussionen. Detta förhållande återspeglas i referatet.

Inledningsvis försökte gruppen få klart för sig var skillnaderna låg mellan responsive och traditionell evaluering. Vi kom då fram till att de viktigaste skillnaderna var att medan man i traditionell evaluering försökte att utgå från på förhand uppställda mer eller mindre preciserade mål och att man dessutom koncentrerade sig på effektsidan, så var syftet med responsive evaluering att inta en mer öppen och flexibel hållning. Detta innebar både en metodisk öppenhet och att målen kunde förändras medan evalueringen pågick. Gruppen menade också att den "nya" metoden innebar att man använde sig av mer kvalitativa data och jämföreleser gjordes med den "mjukdatadebatt" som fördes bland vissa sociologer under slutet av 1960-talet i Sverige. Paralleller sågs också med den av Arne Trankell i Stockholm bedrivna "action cum research". Vidare framhölls det i gruppen att den mindre betoningen på utbildningens effekter i den responsiva evalueringen gjorde att man i större utsträckning uppmärksammande andra relevanta faktorer i undervisningssituationer som t.ex. processvariabler.

Redan så långt stod det klart för gruppen att metoden inte var "ny" i så motto att metodiken inte tillämpats tidigare. Det nya låg snarare i att metoden som prövats i andra sammanhang (t ex socialantropologiska och sociologiska studier av typ Oscar Lewis och Howard Becker) applicerats på ett område där den traditionellt kvantitativa metoden hade haft ett av sina starkaste fästen på grund av att forskarna vanligtvis hade en psykometrisk bakgrund.

En annan faktor som bidragit till att man började pröva nya vägar var de relativa misslyckandena som tidigare effektforskning, t ex vad gäller lärareffektivitet, uppvisat.

Vi kunde alltså komma fram till att Stakes tankegångar passade bra in i den allmänna vetenskapsteoretiska debatt och omorientering som börjat äga rum sedan mitten av 1960-talet och som tagit sig uttryck i en allmän mindre tilltro till vetenskapens förmåga att lösa alla problem samt i diskussioner och debatter som berört objektivitet, skillnader mellan natur- och humanvetenskaper, kvantitativa kontra kvalitativa data, forskarens ansvar etc.

Gruppen diskuterade också olika metodiska problem i anslutning till evaluering. Bland annat diskuterades på vilken systemnivå (mikro-makronivå) den responsiva evalueringen var mest användbar. De flesta i gruppen ansåg att den hade sitt största värde på mikronivån på grund av sin intensiva karaktär. Det innebar dock inte att den var oanvändbar på större system utan endast att man där var tvungen att bara ta ut ett fåtal aspekter, skolor etc. Gymnasieinspektörerna och deras rapporter togs som svenska exempel på sådana försök.

Gruppen kunde också se flera metodiska likheter i uppläggning mellan responsive evaluering och vad som i traditionell metodlitteratur benämnes som pilot- eller förundersökningar och frågan togs därför upp om inte den responsiva evalueringens största nytta låg just i att man fick förhands- och insideinformation om det system som skulle utvärderas. Detta kunde sedan ligga till grund för evaluering av mer traditionell typ. Samtidigt var gruppen emellertid överens om att båda typerna av evaluering behövdes.

What do ideas about responsive evaluation suggest for the training or retraining of educational researchers?

Många egenskaper krävs av en god "responsive" evaluerare, kunskap om de "gamla" psykometriska metoderna, produktmätningar o dyl, men också helgjuten personlighet, god journalistisk förmåga, kapacitet för "portrayal" och att få med relevanta mjukdata i utvärderingen. Frågan är om dessa egenskaper kan förenas i en enda person, kanske team-work är att föredra här. Problem med beskrivning av mjukdata är generaliserbarheten.

Should the evaluator be reluctant to use unsolicited observations and casual opinions as data? Can we validate or corroborate such data?

Vid användning av ovan beskrivna data föreligger ett valideringsproblem. Vi måste veta vad vi är ute att titta efter, annars kan vi "få fel svar på fel frågor". Beroendet av uppdragsgivaren kan bli aktuellt här. Kan vi nöja oss med att klienten är nöjd? En motsättning i Stakes beredskap att tillmötesgå klienterna och hans berättelse (i gruppen) om en uppdragsgivare som ej ville ha publicitet åt sina data. Upplevelsedata är svåra att kommunicera, vi har ju en gång inom psykologin gått ifrån t ex introsepktionen. Men livslängden för "odugliga" metoder (både för klient och "audience") är nog kortare nu än för några årtionden sedan. Skall man kalla denna typ av evaluering för "forskning"? En beskrivande utvärdering sker redan nu i Sverige t ex av Pedagogiska Nämnden, i experimentskolor o dyl. Genomslagskraften hos en ny evalueringsmetod hänger mycket på om den kan bli "rumsren" i akademiska sammanhang, accepteras i en doktorsavhandling, o s v.

Should the methods of evaluation differ for different subject matters, for different agegroups, for different types of students and teachers?

Svaret på denna fråga ansågs vara "ja". Olika modeller kan vara tillämpbara. Mycket av den gamla arsenalen är bra att ha kvar, men vi behöver ett vidare klientbegrepp. Särskilt på postgymnasial nivå kan de som deltar i programmet vara med i utvärderingen, att sätta mål o dyl.

I gruppen diskuterades fördelarna med en beskrivande utvärdering i stil med Trankells "Flisan". Liknar Stakes "situation som man måste lösa". Vissa komponenter i Flisan kan vara generaliserbara, de kan kanske också lättare meddelas i denna form till t ex kommunalgubbar o dyl. i beslutande ställning. Den psykometriska metoden har blivit prestigeladdad.

DELTAGARLISTA till Symposiet NEW TRENDS IN EVALUATION:

Namn	Institution/Motsv.	Ort
Aldemark, Lars	Ped. inst.	Göteborg
Andersson, Bengt Erik	Ped. inst.	Göteborg
Andersson, Lillemor	Psyk. inst.	Göteborg
André, Raoul	Ped. inst.	Göteborg
Askling, Berit	Lärarhögskolan	Kalmar
Axelsson, Rune	Lärarhögskolan	Uppsala
Barry, David F	Amerikanska Ambassaden	Stockholm
Berggren, Ingegerd	Ped. inst.	Uppsala
Björklund, Eskil	UKÄ:s FoU-enhet	Stockholm
Björsjö, Märta	Göteborgs Skolförvaltning	Göteborg
Bladini, Ulla-Britt	Lärarhögskolan	Göteborg
Bäck, Per	Ped. inst.	Göteborg
Carlsten, Yngve	Lärarhögskolan	Stockholm
Colliander, Marianne	Ped. inst.	Göteborg
Dahlgren, Lars Owe	Ped. inst.	Göteborg
Darås-Olsson, Kerstin	Lärarhögskolan	Malmö
Drambo, Leif	UKÄ/U	Stockholm
Ekholm, Mats	Ped. inst.	Göteborg
Elison, Åke	Ped. inst.	Göteborg
Ellmin, Roger	Psyk. inst.	Stockholm
Ellström, Per-Erik	Ped. inst.	Linköping
Ericson, Lars	Ped. inst.	Göteborg
Gestrelius, Kurt	Lärarhögskolan	Malmö
Graham, Peter	Tekn. högskolan	Stockholm
Gustafsson, Bertil	Lärarhögskolan	Göteborg
Gustafsson, Christina	Ped. inst.	Göteborg
Gustafsson, Jan-Eric	Ped. inst.	Göteborg
Handal, Gunnar	Ped. inst.	Linköping
Holmström, Lars Gunnar	Roskilde Universitet	Roskilde/Danmark

Namn	Institution/Motsv.	Ort
Höghielm, Robert	SÖ, sektion L3:3	Stockholm
Idman, Pekka	Lärarhögskolan	Malmö
Jansson, Bengt-Eric	Psyk. inst.	Stockholm
Janson, Sven	Ped. inst.	Umeå
Johansson, Gunnel	Ped. inst.	Göteborg
Johansson, Kjell	Vuxengymnasium III	Göteborg
Johansson, Martin	Ped. inst.	Umeå
Jonsson, Maud	Lärarhögskolan	Uppsala
Kallós, Daniel	Ped. inst.	Lund
Klingberg, Göte	Lärarhögskolan	Göteborg
Kärrby, Gunni	Ped. inst.	Göteborg
Köhler, Eva-Marie	Ped. inst.	Lund
Lepik, Andrus	Ped. inst.	Göteborg
Levén, Sören	Skolöverstyrelsen	Stockholm
Levin, Lennart	Lärarhögskolan	Göteborg
Lindberg, Maj-Britt	Ped. inst.	Umeå
Lindblad, Sverker	Ped. inst.	Göteborg
Lindström, Berner	Ped. inst.	Göteborg
Lundman, Lars	Lärarhögskolan	Stockholm
Lövgren, Esse	Skolöverstyrelsen L3:1	Stockholm
Lövgren, Horst	Ped. inst.	Lund
Magne, Olof	Lärarhögskolan	Lund
Markesjö, Gunnar	Tekn. Högskolan	Stockholm
Marton, Ference	Ped. inst.	Göteborg
Nilsson, Lennart H S	Ped. inst.	Göteborg
Nordén, Bengt	Ped. inst.	Göteborg
Nordström, Anna-Lena	Lärarhögskolan	Malmö
Nyström, Astrid	Skolöverstyrelsen	Stockholm
Olsson, Ingela	PA-Rådet	Göteborg
Patriksson, Göran	Ped. inst.	Göteborg
Petersson, Lennart	Ped. inst.	Stockholm

Namn	Institution/Motsv.	Ort
Rolf, Magnus	Ped. inst.	Göteborg
Rosén, Kerstin	Ped. inst.	Göteborg
Rovio-Johansson, Airi	Ped. inst.	Göteborg
Rubenowitz, Sigvard	Psyk. inst.	Göteborg
Rudemo, Lisbeth	Skolöverstyrelsen, Byrå L3:3	Stockholm
Sandgren, Björn	Ped. inst.	Göteborg
Schulenburg, Rosvita	Esselte	Stockholm
Silverberg, Bengt	Skolöverstyrelsen	Stockholm
Solstad, Arne	Forsöksrådet för Skoleverket	Oslo
Stake, Bernadine		Göteborg
Stencrantz, Annika	Ped. inst.	Göteborg
Stukát, Karl-Gustaf	Lärarhögskolan	Göteborg
Svanberg, Gunilla	Ped. inst.	Göteborg
Svensson, Harry	Lärarhögskolan	Uppsala
Svingby, Gunilla	Ped. inst.	Göteborg
Säljö, Roger	Ped. inst.	Göteborg
Thång, Per Olof	Ped. inst.	Göteborg
Wenestam, Claes-Göran	Ped. inst.	Göteborg
Westrin, Per Anders	Ped. inst.	Lund
Wicknertz, Anne-Marie	Ped. inst.	Göteborg
Ölund, Ulf	Ped. inst.	Umeå

REFERENSER

American Psychological Association, 1966: Standards for Educational and Psychological Testa and Manuals, Revised. Washington, D.C.: American Psychological Association.

Anderson, R.C. 1969: "The comparative field experiment: an illustration from high school biology." Proceedings of the 1968 Invitational Conference on Testing Problems. Princeton, New Yersey: Educational Testing Service, p.p. 3-30.

Astin, A. and Panos, R. 1971: The Evaluation of Educational Programs. In Thorndike, R. (ed.) Educational Measurement, 2nd ed. Washington, D.C.: American Council on Education.

Coleman, J. et al. 1968: Equality of Educational Opportunity. Washington, D.C.: Government Printing Office.

Cronbach, L.J. 1963: "Course improvement through evaluation". Teachers College Record, 64, pp 672-683.

Eisner, E. 1972: "Emerging models for educational evaluation." School Review, August, pp 573-589.

Flexner, A. 1910: Medical Education in the United States and Canada: A Report to the Carnegie Foundation for the Advancement of Teaching. Boston: D.B. Updike, The Merrymount Press.

Heath, R.W. (ed.) 1964: New Curricula. New York: Harper o. Row.

House, E.R. 1973: "The conscience of educational evaluation". In House, E.R. (ed.) School evaluation, the politics and process. Berkely, California: McCutchan Publising Co.

House, E.R. 1973: "The dominion of economic accountability". In House, E.R. (ed.) School evaluation, the politics and process. Berkeley, Californa: McCutchan Publishing Co.

Husén, T. (ed.) 1967: International study of achievement in mathematics. New York: Wiley ans Sons Inc., and Stockholm: Almqvist o. Wiksell.

Härnqvist, K. 1974: The international study of educational achievement. Manuscript.

Komoski, K. 1967: See Stake, R.E. 1967.

Kourilsky, M. 1973: "An adversay model for evaluation". Paper presented at the annual meeting of the American Educational Research Association, New Orleans, February 26, 1973.

Lindquist, E.F. (ed.) 1951: Educational Measurement. Washington, D.C.: American Council on Education.

MacDonald, B. 1973: "Briefing decision-makers - the evaluation of the Humanities Curriculum Project." In House, E.R. (ed.) Educational Evaluation, the politics and process. Berkeley, California: McCutchan Publishing Co.

National Study of Secondary School Evaluation, 1970: Evaluative Criteria, Fourth Edition. National Study of School Evaluation, 2201 Wilson Blrd, Arlington, Va.

Owens, T.R. 1973: Educational Evaluation by Adversay proceeding," In House, E.R. (ed.) School Evaluation, the politics and process. Berkeley, Californa: McCutchan Publishing Co.

Parlett, M. and Hamilton, D. 1972: "Evaluation as illumination: A new approach to the study of innovatory programs". Accasional paper No 9, Centre for Research in the Educational Sciences, University of Edinburgh.

Popham, W.J. 1969: "Objectives and instruction". AERA monograph Series on Curriculum Evaluation, Vol. 3, Chicago: Rand Mc Nally Co., pp 32-52.

Provus, M. 1972: "The discrepancy evaluation model". In Taylor, P.A. and Crowley, D.M. (eds.). Readings in Curriculum Evaluation. Dubuque, Iowa: William C. Brown, pp 117-127.

Rippey, R.M. 1973: Studies in Transactional Evaluation. Berkeley, California: Mc Cutchan Publishing Co.

Scriven, M. 1967: The methodology of evaluation". AERA Monograph Series on Curriculum Evaluation. Vol. 1. Chicago: Rand McNally Co.

Scriven, M. 1972: "Prose and Cons about Goal-Free Evaluation". Evaluation Comment, December 1972, Vol. 3, No 4. Center for the Study of Evaluation, UCLA, Los Angeles, California.

Smith, L.M. 1972: "The use of observational strategies for the development of a summative evaluation model". Paper presented at the annual meeting of the American Educational Research Association. Chicago, April 7, 1972.

Smith, L.M. and Pohland, P.A. In Press: "Education, Technology and the Rural Highlands". In AERA Monograph Series on Curriculum Evaluation. Vol. 7. Chicago: Rand McNally Co.

Smith, E.R. and Tyler, R.W. 1942: Appraising and recording student progress. New York: Harper and Row.

Stake, R.E. (ed.) American Educational Research Association Monograph Series on Curriculum Evaluation. Chicago: Rand McNally Co.

Stake, R.E. 1967: "A research rationale for EPIE". The EPIE Forum. Vol. 1, No 1, September 1967.

Stake, R.E. 1967: "The countenance of Educational Evaluation". Teachers College Record, 68, pp 523-540.

Stake, R.E. and Gjerde, C. In Press: "An evaluation of TCITY. The Twin City Institute for Talented Youth". In AERA Monograph Series on Curriculum Evaluation. Vol. 7. Chicago: Rand McNally.

Stufflebeam, D.L. 1969: "Evaluation as enlightenment for decision-making". In Beatty, W.H. (ed.): Improving Educational Assessment and An Inventory Affective Behavior. Washington: Association for Supervision and Curriculum Development. National Educational Association, pp 41-73.

Stufflebeam, D.L. et al. 1971: Educational Evaluation and Decision-Making. Itasca, Illinois: Peacock Publishers.

Taba, H. 1966: "Teaching strategies and cognitive functioning in elementary school children". Cooperative Research Project No 2404. San Francisco State College, San Francisco.

Tyler, R.W. 1942: "General statement on evaluation". Journal of Educational Research, March 1942, pp 492 - 501. Also see Smith and Tyler.

Tyler, R.W. 1966: "The objectives and plans for a National Assessment of Educational Progress. Journal of Educational Measurement. 3, pp 1 - 4.